



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사학위논문

서지결합분석을 활용한 국내 학문
영역 간 유사 관계 도출 및 시각화

Analyze and visualize the relationship of
domestic academic disciplines using
bibliographic coupling

2014 년 2 월

서울대학교 대학원
융합과학부 디지털정보융합전공
박 기 범

초록

본 연구에서는 인용 관계 분석 기법의 하나인 서지결합분석을 활용하여 국내의 학문 영역 간 유사 관계를 도출하고, 도출된 결과를 시각화하여 광역 과학 지도를 생성하였다. 학문 영역 간 유사 관계 도출 및 시각화를 위해, 국내의 다양한 학술지에 대한 등재 문헌 정보 및 인용 정보를 제공하고 있는 한국학술지인용색인(Korea Citation Index, KCI)의 인용 데이터베이스를 활용하였다. 학문 영역 간 유사 관계 도출에는 빈도 행렬에 기반을 둔 단순 방법, 최소 방법, 그리고 혼합 방법을 활용하거나, 유사성 행렬에 기반을 둔 자카드 유사 계수, 코사인 유사도, 그리고 피어슨 상관계수를 활용하였다. 다음으로, 정의된 유사 관계 설정 방법에 따른 학문 영역 간 유사 지수 행렬을 도출하고, 각각의 도출 방법에 따른 시각화 결과를 비교하여 공통점 및 차이점을 분석하였다. 학문 영역 간 관계의 시각화 과정에는 패스파인더 네트워크 알고리즘을 적용하여 유사성이 높은 학문 영역 간 연결이 명확하게 드러날 수 있도록 하였다.

위의 일련의 과정을 통하여, 지난 10 년간 활발히 연구되었던 국내 주요 학문 영역 간 유사 관계를 다양한 방법을 활용하여 계량적으로 도출하였다. 또한, 각 방법으로 도출된 결과를 바탕으로 학문 영역 간 유사 관계를 직관적으로 파악할 수 있는 광역 과학 지도를 생성하였다. 최종적으로 생성된 지도들에 대한 비교 분석 및 평가를 통하여 학문 영역 간 관계 시각화에 적절한 유사 지수 도출 방법을 실험적으로 확인하였다.

주요어: 서지결합분석, 과학지도학, 패스파인더 네트워크, 정보 시각화,
한국학술지인용색인, KCI
학번: 2012-22463

목차

제 1 장 서론	1
1.1 연구 배경	1
1.2 연구 목적	2
제 2 장 이론적 배경	4
2.1 인용분석과 동시인용분석	4
2.2 저자동시인용분석	6
2.2.1 정의 및 특징	6
2.2.2 선행 연구	8
2.3 저자서지결합분석	11
2.3.1 정의 및 특징	11
2.3.2 선행 연구	16
2.4 네트워크 분석 기법	19
2.4.1 다차원척도법	20
2.4.2 패스파인더 네트워크	22
2.5 과학지도학	26
2.5.1 정의 및 특징	26
2.5.2 선행 연구	26
제 3 장 연구 문제	29

제 4 장 연구 방법	31
4.1 연구 대상 선정	31
4.2 데이터 수집	31
4.3 데이터 사전 처리	32
4.4 서지결합분석 수행	33
4.5 관계 시각화	35
제 5 장 결과 및 논의	36
5.1 빈도 지수 기반 관계 분석	36
5.1.1 단순 방법 활용 결과	36
5.1.2 최소 방법 활용 결과	39
5.1.3 혼합 방법 활용 결과	41
5.1.4 빈도 지수 기반 결과의 상호 비교	44
5.2 유사도 지수 기반 관계 분석	46
5.2.1 자카드 유사 계수 활용 결과	46
5.2.2 코사인 유사도 활용 결과	50
5.2.3 피어슨 상관계수 활용 결과	53
5.2.4 유사도 지수 기반 결과의 상호 비교	57
5.3 활용 지수에 따른 시각화 결과 비교	58
5.4 연구의 한계	59
5.4.1 동일 참고문헌의 판별 문제	60
5.4.2 KCI 참고문헌 데이터의 미비	61

5.4.3 도출된 관계에 대한 평가의 어려움	62
제 6 장 결론	63
참고문헌	65

표 목차

표 1. MA 의 유사도 도출 방법별 유사 지수 예시	15
표 2. 단순 방법을 활용한 학문 분야 간 유사 지수 (상위 10 항목)	36
표 3. 단순 방법을 활용한 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 (상위 10 항목)	37
표 4. 최소 방법을 활용한 학문 분야 간 유사 지수 (상위 10 항목)	39
표 5. 최소 방법을 활용한 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 (상위 10 항목)	40
표 6. 혼합 방법을 활용한 학문 분야 간 유사 지수 (상위 10 항목)	42
표 7. 혼합 방법을 활용한 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 (상위 10 항목)	43
표 8. 자카드 유사 계수를 활용한 학문 분야 간 유사 지수 (상위 10 항목)	47
표 9. 자카드 유사 계수를 활용한 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 (상위 10 항목)	48
표 10. 코사인 유사도를 활용한 학문 분야 간 유사 지수 (상위 10 항목)	50
표 11. 코사인 유사도를 활용한 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 (상위 10 항목)	52

표 12. 피어슨 상관계수를 활용한 학문 분야 간 유사 지수 (상위 10 항목)	54
표 13. 피어슨 상관계수를 활용한 상위 대분류 영역이 서로 다른 학문 분야 간 관계 지수 (상위 10 항목)	55

그림 목차

그림 1. 단순 방법을 활용한 학문 분야 간 패스파인더 네트워크	38
그림 2. 최소 방법을 활용한 학문 분야 간 패스파인더 네트워크	41
그림 3. 혼합 방법을 활용한 학문 분야 간 패스파인더 네트워크	44
그림 4. 자카드 유사 계수를 활용한 학문 분야 간 패스파인더 네트워크	49
그림 5. 코사인 유사도를 활용한 학문 분야 간 패스파인더 네트워크	53
그림 6. 피어슨 상관계수를 활용한 학문 분야 간 패스파인더 네트워크	56

제 1 장 서론

1.1 연구 배경

사회 구조가 복잡해지고 다양해짐에 따라, 사회에서 발생하는 여러 문제 또한 다양한 요소들과 연결되어 이를 단일 학문 영역 측면에서의 접근을 통해 해결하는 것이 점차 어려워지고 있다. 이를 해결하기 위해, 문제에 대하여 다양한 학문 영역의 접근 방법을 활용하는 이른바 문제 중심형 연구가 주요한 연구 방법으로 떠오르고 있으며, 문제 해결 과정에서는 서로 다른 학문 영역에 속한 연구자들 간 협업의 필요성이 커지고 있다. 이러한 흐름 속에서 학문 간 교류의 시도가 지속해서 증가하고 있으며, 학문 영역 간 관계에 대한 관심 또한 증대되고 있다.

한편, 정보과학에서는 분석 대상 간 관계를 도출하고 지적 구조를 파악하기 위한 연구를 지속해서 수행해오고 있다. 특히 자료의 전산화를 통하여 대용량의 정보 처리가 가능해짐에 따라, 대량의 자료로부터 가치 있는 정보들을 추출하여 이를 활용하고자 하는 방법들이 활발하게 연구되고 있다. 이러한 방법들은 비단 새로운 창출되는 영역에서뿐만 아니라, 기존에는 여러 현실적 제약으로 시도하지 못하였던 대단위 문헌 분석과 같은 전통적인 연구 영역의 과제들에도 많이 적용되고 있다.

문헌 분석의 기본 대상이자 문헌정보학의 주요 연구 자료 중 하나인 학술 문헌은 연구자들이 자신의 연구 결과를 발표하기 위하여 생성한 문서이다. 학술 문헌에는 저자의 정보와 저자가 연구활동에서 참고한 문헌들의 정보가 명확하게 표현되는 형태적 특징을 가지고 있다. 이러한 형태적 특징을 활용하여, 문헌의 참고문헌 관계를 바탕으로

문헌, 저자 또는 학술지 간 유사 관계를 도출하는 인용 관계 분석 연구가 다양한 학문 영역을 대상으로 이루어지고 있다. 더 나아가, 도출된 유사 관계를 시각화하여 추가적인 유용한 정보를 얻고자 하는 이른바 과학지도학(Scientography)이 시작되어 점차 그 적용 영역을 확대해 나가고 있다.

국내에서도 문헌의 인용 관계를 분석하는 다양한 연구가 진행되고 있었으나, 그동안 국내 문헌에 대한 양질의 정보를 제공하는 인용 데이터베이스가 구축되어 있지 않아 그 연구의 한계가 있었다. 그러나 2005 년경 한국학술지인용색인(Korea Citation Index, KCI) 데이터베이스가 구축되어 현재까지 약 10 여 년의 자료 축적을 통하여 총 93 만 건 이상의 문헌과 1 천 3 백만 건 이상의 참고문헌 정보를 보유하게 되었다. 따라서 인용 데이터베이스의 축적된 정보를 활용한 인용 관계 분석 연구의 가능성이 매우 커지게 되었다.

1.2 연구 목적

본 연구에서는 학술 문헌의 인용 관계를 활용하여 국내의 학문 영역 간 유사 관계를 도출하고 이를 시각화하는 것을 목표로 한다. 분석의 대상으로는 국내의 다양한 학문 영역의 학술지에 대한 등재 문헌 및 인용 정보를 제공하고 있는 KCI 데이터베이스를 활용한다. 학문 영역 간 유사 관계 정의를 위한 방법으로는 서지결합방법을 활용하여 공통 참고문헌의 존재 및 인용 횟수에 따라 유사 관계가 정의될 수 있도록 한다. 유사 지수는 빈도 기반 지수 3 개와 유사도 기반 지수 3 개 등 총 6 개의 지수를 도출하고 그 결과를 비교한다. 또한, 도출된 유사 지수를 활용하여 학문 영역 간 관계를 시각적으로 표현하고 이를 비교하여 공통점 및 차이점을 찾는다. 이때 시각화 과정에서 패스파인더

네트워크 알고리즘을 적용하여 유사성이 높은 학문 영역 간 관계가 명확하게 드러날 수 있도록 한다. 마지막으로 유사 지수별 시각화 결과를 바탕으로 관계 시각화에 적절한 유사 지수 도출 방법을 실험적으로 확인한다. 위의 일련의 과정을 통하여, 동시인용분석보다 상대적으로 수행된 연구가 적은 서지결합분석 방법을 활용하여, 국내 인용 데이터베이스를 바탕으로 학문 영역 간 유사 관계를 확인하고, 유사 지수 도출 방법에 따른 결과의 차이 및 시각화への 적합성 여부를 판단하며, 국내의 주요 학문 영역을 포괄하는 광역 과학 지도를 생성하는 데 본 연구의 의의가 있다.

제 2 장 이론적 배경

2.1 인용분석과 동시인용분석

인용 분석은 “누가 누구를 인용하는지”의 기록 정보를 활용한 관계 분석의 형태 중 하나이다 [1]. 인용의 양식에 관한 연구는 정보과학 분야에서 가장 많이 논의되는 연구 주제 중 하나이며 [2], 인용은 문헌과 그 문헌의 저자, 저자의 소속기관 등 다양한 분석 단위들 사이의 관계를 형성할 수 있다.

인용 분석이 함께 인용된 문헌과 그 문헌의 저자, 저자의 소속기관 등의 분석 단위를 포함할 때 이를 동시인용분석이라고 하며, 이는 현재 가장 널리 쓰이고 있는 계량적 분석 방법의 하나이다. 동시인용분석을 제안한 Small 은 동시인용을 “이전에 생성된 문헌 중 두 개가 이후에 생성되는 문헌에 의해 같이 인용되는 빈도”로 정의하였다 [3]. 만약 어떤 문헌들이 다른 문헌들에 비하여 더 빈번하게 동시 인용되면, 그 문헌들은 다른 문헌들에 비해 서로에게 더 많이 관계가 있는 것으로 간주한다. 또한, 일정한 기간 동시인용분석으로 도출되는 관계의 변화를 관찰하면 학문 영역의 변화와 지식의 성장을 확인할 수 있다고 하였다 [4].

Small 과 Griffith 는 빈번히 동시 인용되는 문헌들의 군집을 분석하여 학문의 흐름의 변화에 영향을 준 학자들의 군집을 판별하는 방법을 제안하였다 [5]. 그들은 동시 인용된 문헌들의 군집 분석을 통해 그 문헌들의 저자들이 분석 대상으로서의 가치가 있다는 것을 인지했다. 이후 Small 은 동시인용 정보를 학문적 “전공(specialty)”의 변화를 판별하는 방법으로 사용하였다 [6]. 여기서 “전공”은 “현대 학문의 사회적이고 인지적인 조직의 주된 형태(the principal mode of social

and cognitive organization in modern science)”를 뜻하며, 빈번히 동시 인용되는 문헌들의 군집은 그 분야의 인지적인 구조를 대표한다고 여겨진다. 이러한 군집들의 공간을 “전공의 현재 체계(current paradigm of the specialty)”라고 하며 군집에 있는 문헌의 저자들을 “엘리트(elite)” 또는 “전공의 선도 연구자(leading scientist of the specialty)”라고 하였다.

Small 에 따르면 동시인용분석은 대상 연구분야에서 빈번히 인용된 문헌들의 군집을 포함하며, 이 분석 방법은 “학문의 포괄적인 지도를 구성하는 데 중요한 단계(a key step in constructing comprehensive maps of science)”라고 하였다 [7]. 또한, Griffith 는 동시인용분석이 “학문의 구분과 변화의 속도와 규모에 대한 실제 모습(a realistic picture of differentiation in science and the speed and extensiveness of change)”을 생성한다는 것을 강조하였다. 즉, 동시인용분석은 대상 연구분야의 지적 구조를 판단하는 것뿐만 아니라 그 구조의 변화를 관찰하는 데도 활용될 수 있다.

인용분석은 문헌의 인용 빈도를 통해 분석 대상 영역의 핵심 문헌을 판별할 수 있다. 나아가, 빈번히 인용된 문헌은 다른 문헌들보다 학자들에게 더 영향력 있는 핵심 문헌으로 판단될 수 있다. 한편, 인용분석에서의 핵심 문헌들은 네트워크 분석에서는 중심 교점(node)들로 표현될 수 있다. 또한, 빈번히 인용된 문헌이나 저자는 더 높은 연결 강도를 가지고 있는 것으로 간주할 수 있으므로, 인용의 횟수는 연결 강도로 활용될 수 있다. 연결의 강도는 “시간, 감정적 강렬함, 친밀함(상호 간의 신뢰), 그리고 묶음의 특징이 되는 상호 도움의 양에 대한 (아마도 선형의) 조합(a (probably linear) combination of the amount of time, the emotional intensity, the

intimacy(mutual confiding), and the reciprocal services which characterizes the tie)”을 의미할 수 있다 [8]. 따라서 연결 강도의 개념을 인용 분석에 적용하는 것이 가능하다.

2.2 저자동시인용분석

2.2.1 정의 및 특징

지적 구조를 분석하기 위한 최근의 주요 연구 방법은 인용 정보에 기반을 둔 저자 관계 분석이다. 이 분석은 저자 간 관계성을 판단하기 위해 분석 대상 저자들의 문헌들과 그에 포함된 인용 정보에 대하여 동시인용 빈도 등의 인용 기반 측정을 통해 관계를 도출하고 지적 구조를 관찰한다. 학술 문헌에서 인용된 문헌은 인용한 문헌의 이론적, 실증적 토대가 무엇인지를 나타내므로, 인용된 문헌을 분석해보면 같은 학문적 배경을 가지는 문헌 또는 저자 간 관계를 파악할 수 있다 [9].

White 와 Griffith 가 제안한 저자동시인용(author cocitation analysis, ACA)은 어떤 저자가 자신의 문헌에 이전에 발표된 다른 저자의 문헌과 또 다른 저자의 문헌을 동시에 인용하는 경우 발생한다 [10]. 다만 동시인용분석과는 달리 어떠한 문헌이 인용되었는지는 관계가 없고, 특정한 저자에 의하여 두 명의 저자가 동시에 인용되었는지를 활용하여 관계를 도출한다. 따라서 분석의 기본 단위는 저자가 되며, 그 저자가 쓴 저작물 전체가 분석의 기반 정보가 된다.

저자동시인용은 특정 저자가 다른 저자를 인용하였다면 다른 저자는 특정 저자에게 영향을 주었다고 판단할 수 있으며, 특정 저자가 다른 저자와 또 다른 저자를 동시에 인용하였다면 다른 저자와 또 다른 저자는 서로 연관되어 있다는 것을 전제로 한다. 이는 특정 저자의

문헌이 어느 선행 저자의 영향을 받았는가를 의미하는 동시에, 빈번히
동시 인용되는 저자들은 서로 밀접한 관계가 있음을 의미한다. 따라서
저자동시인용분석은 학문의 여러 세부 주제 영역에 대한 지적 구조를
도출하고 분석하기에 유용한 방법이다.

이러한 저자동시인용분석의 개념과 방법은 White 와 Griffith 가
사회학 분야의 대표 학자를 대상으로 한 연구 결과에서 그 타당성이
입증되었는데 [10], 그들은 특정 저자 묶음이 동시 인용된 문헌들이
서로 밀접한 관계가 있으며, 각 저자 묶음이 그들의 저작물에 담긴
지식 구조를 대표할 수 있다는 것을 증명하였다. 즉, 단일 인용된
저자만으로는 인용한 문헌의 주제에 대한 함축적 의미가 모호하지만,
묶음으로 인용된 저자는 주제에 대한 높은 함축적 의미를 가진다는
것이다 [11].

이러한 저자동시인용과 분석 대상 저자들 간의 관계를 활용하여,
저자동시인용은 동시 인용된 저자들의 지도를 생성하여 분석 대상
영역의 지적 구조를 보여주기 위해 활용되었다 [4], [10], [12]. 또한,
저자동시인용분석 방법은 특정 주제 영역의 변화를 시간에 걸쳐
분석하거나, 전통적인 세부 연구 주제 간 융합의 가능성 등을 검증하기
위하여 사용되었다 [13].

저자동시인용분석은 문헌이 아닌 저자를 분석 대상으로 하여 저자가
특정 주제를 포괄적으로 나타내는 특성을 활용하기 때문에, 연구자가
분석 대상 영역의 상세한 문헌 정보와 전문 지식을 가지고 있지
않더라도 분석이 가능한 장점이 있다 [14], [15]. 그러나
저자동시인용분석은 문헌이 다른 사람에게 인용되기까지 필요한
문헌의 발행주기에 따르는 수년의 통상적 지체 현상이 존재한다.
따라서 저자동시인용분석은 수년 전 과거에 문헌을 발행한 저자들의

정보를 활용하여 분석하게 되므로 과거의 지적 구조를 관찰하는 방법이 되며, 수년 전의 지적 구조를 분석할 수는 있으나 현재의 연구 경향과 학술 활동을 활발히 수행하는 연구자를 파악하기 어렵다는 단점이 있다 [16], [17]. 이러한 한계 상황은 예시적으로 (1) 저명한 연구자가 이미 은퇴하였거나 고인이 되었음에도 지속해서 인용되는 경우, (2) 연구자의 연구 분야가 변화함에 따라서 특정 시기에 인용되는 문헌의 주제와 발표한 문헌의 주제가 일치하지 않는 경우, (3) 국내 연구 동향을 분석할 때 특정 주제와 관련하여 인용되는 저자와 문헌을 발표하는 저자가 다른 경우 등이 지적되었다 [16].

한편, 저자를 분석 단위로 하고 동시 인용된 빈도수를 각 저자 간 거리를 나타내는 변수로 사용하여 저자 간 관계를 지도로 표현하는 저자동시인용지도(author cocitation map)는 분석 대상 학문 영역의 지적 구조를 시각적으로 표현해 준다 [18]–[20]. 지도에서 동시인용 빈도가 높은 저자들은 서로 가깝게 나타나며, 빈도가 낮은 저자들은 서로 비교적 멀리 나타나는데, 이 위치는 수많은 인용자들의 복합적인 판단으로 결정되어 나타나는 것이므로 매우 객관적인 분석이라 할 수 있다. 이러한 표현 방법은 매우 큰 규모의 데이터를 사용하여 오랜 기간에 걸쳐 반복적으로 나타나는 패턴을 활용한 것이므로, 이를 통해 생성된 지도는 대상 영역의 한 사람의 학자가 분석하기 어려운 학문의 지적 구조를 거시적 관점에서 분석할 수 있게 한다 [21].

2.2.2 선행 연구

저자동시인용분석은 학문 영역의 지적 구조를 분석하고 연구 결과를 계량적으로 측정하여 연구의 흐름과 학문의 발전 방향을 파악할 수 있게 하였다. 저자동시인용분석은 이론적 측면의 많은 연구와 정교화

과정을 통하여 대상 간 관계를 분석하는 한 방법으로 정립되었으며 [12], [22], [23], 철학 [24], 의료정보학 [25], 기록관리학 [26], 정보학 [27], 연구 정책 [28] 등 다양한 학문 영역을 대상으로 하는 연구들이 지속해서 수행되었다.

특히 1990 년 Journal of American Society for Information Science 에서 저자동시인용분석을 특집 주제로 선정한 것은 저자동시인용분석을 널리 확산시키는 계기가 되었다 [29]. 이 특집에서 발표된 McCain 의 개관 논문과, 단행본에 수록된 White 의 개관 논문은 저자동시인용분석이 정립되고 확산하는 데 큰 도움을 주었다 [15], [20].

Rice 등은 1977 년부터 1985 년까지의 커뮤니케이션 저널 간 인용 관계에 관한 연구에서, 인용 정보의 서지 분석과 인용 패턴에 대한 관계 분석을 통해 대인관계에 관련된 저널과 대중매체 저널, 그리고 지역적으로 고립된 저널의 군집을 판별하였다 [30]. 서지 분석은 커뮤니케이션 분야의 핵심 저널을 판별하기 위해 사용되었으며, 관계 분석은 응집 분석과 위치 분석을 통해 인용 패턴과 저널 군집을 판별하기 위해 사용되었다. 응집 분석은 “교점 간 연결 정도(the degree of linkages between any pair of nodes)”에 기반을 두었으며 위치 분석은 “교점 간 구조적 동등성(the structural equivalence between nodes)”에 기초하였다. 위치 접근에서는 다른 저널을 인용하거나 다른 저널로부터 인용 받는 것 모두의 상관성에 기반을 두고, 핵심 저널들의 군집을 판별하기 위해 군집 분석과 다차원척도법을 사용하였다. 그들은 연구 결과를 통해, 학문 정보의 흐름에는 다른 중요한 채널이 많이 있다는 점을 고려하더라도, 인용

분석이 학문적 커뮤니케이션의 패턴과 학문 영역의 지적 구조를 분석하는데 매우 유용함을 증명하였다 [30].

McCain 은 저자동시인용분석을 활용하여 거시경제 영역의 지적 구조를 규명하는 연구를 수행하고, 다차원 척도와 군집분석을 통하여 2 차원의 공간상에 표현하였으며 [14], 기간별 분석을 통하여 지적 구조의 변화 및 새로운 학과의 등장과 성장을 보여주었다 [18]. 이외에도 저자동시인용분석을 통하여 도출한 특정 학문 영역의 지적 구조 분석 결과를 해당 영역 전문가와 검증하여 저자동시인용분석이 학문 영역의 지적 구조를 분석하는 데 유용한 방법임을 확인하였다 [19]. 또한, McCain 은 초파리 유전학 분야의 58 명의 저자를 대상으로 1981 년부터 1986 년까지의 인용 정보를 분석하였다 [31]. 분석을 통하여 도출된 지도는 비슷한 학문 주제를 포함하고 있는 저자들의 군집을 보여주었으며, 학문 영역에서의 저자의 기여가 어떠한 측면에서 나타나는지 확인할 수 있게 하였다.

Paisley 는 저자동시인용분석을 통하여 학제적 성격을 가지는 학문을 판단하고자 하였다 [32]. 그는 연구의 결과로 심리학, 사회학, 정치학, 인류학 등의 학문이 커뮤니케이션 연구 분야에 광범위한 영향을 미치고 있음을 보였다.

다양한 분석 기법들 또한 저자동시인용분석에 활용되고 있다. Nerur 등은 1980 년부터 2000 년까지 경영학의 한 분야인 전략경영 분야의 저자들을 대상으로 동시인용분석을 수행하는 데 있어서 다차원척도법, 요인분석, 패스파인더 분석 등을 활용하였다 [33]. 이러한 분석 방법들을 통하여 전략경영 분야의 지적 구조를 2 차원의 공간상에 표현하고, 각 하위 주제 영역 간 관계를 정의하였다. 또한, 연구의

개념적인 연결에 중요한 역할을 한 저자를 찾고, 주제 사이의 공간적 거리를 2 차원 공간상에 시각화하였다.

국내에서도 저자동시인용분석을 사용하여 학문 영역의 지적 구조를 분석하기 위한 다양한 연구들이 수행되었다. 김영진은 사회학 분야의 45 명의 저자를 대상으로 저자동시인용분석을 실시하였다 [34]. 그는 연구를 통하여, 저자들 간 상호인용이 저조한 국내 영역에 대한 상대적으로 낮은 동시인용빈도에 기반을 둔 분석도 의미 있는 결과를 도출할 수 있음을 확인하였다. 민윤경은 국내 경영학 학술지를 대상으로 저자 42 명을 선정하여 저자동시인용분석을 실시하였다 [35]. 그는 연구를 통하여, 저자동시인용분석이 전문가의 도움 없이도 특정 주제 영역에 대한 개괄적인 지적 구조를 파악할 수 있는 유용한 방법임을 확인하였다. 김도미는 1977 년부터 1991 년까지 68 명의 경제학자를 대상으로 저자동시인용분석을 실시하였고, 이를 인용 문헌의 색인어 분석 결과와 비교하였다 [36]. 그는 연구를 통하여 저자동시인용분석이 지적 구조를 분석하는 데 유용한 방법임을 확인하였다.

2.3 저자서지결합분석

2.3.1 정의 및 특징

앞서 설명한 저자동시인용분석이 지적 구조를 분석하는 데 유용하지만, 저자동시인용분석의 여러 한계점으로 인하여 학문의 지적 구조를 분석하기 위한 다른 측면의 방법으로 (저자)서지결합분석이 활용되고 있다.

서지결합(bibliographic coupling)은 Kessler 가 1961 년 MIT report 에 처음으로 제안하고, 1963 년에 발표한 이론이다 [37]. 서지결합분석은 1973 년 Small 에 의해 제안된 동시인용분석보다 10 년 앞서 소개되었으며, 동시인용분석과 그 분석 방향의 측면에서 반대되는 개념이다. [3]. 서지결합은 두 문헌의 인용 목록에 공통되는 문헌들이 더 많을수록, 그 두 문헌의 주제가 비슷할 가능성이 높다는 기본 가정에 기반을 두고 있다. 이때 두 문헌은 “결합된 문헌”이라고 하며, 그것들 사이의 연관성을 서지 결합이라고 한다.

두 문헌은 하나 또는 다수의 문헌을 그들의 인용 목록에 공유할 수 있고, 문헌들이 공유하는 인용 문헌의 수는 문헌 간 주제 유사성의 측정을 위하여 활용될 수 있으며, 이는 결합 강도처럼 사용될 수 있다. 따라서 결합 강도는 두 비교 대상 문헌이 공유하는 인용 문헌의 수로 나타낼 수 있으며, 두 문헌 간 결합 강도가 더 클수록 두 문헌은 더 유사하다고 판단된다.

서지결합은 여러 영역에서 연구 선구자를 탐험하고, 영역의 지적 구조를 발견하고 정보를 수집하기 위해 도입되었다 [38]. 서지결합 빈도에 기반을 둔 지적 구조 분석은 과거의 출판물을 활용하는 동시인용분석에 비하여 더 최근의 출판물에서 연구자들이 문헌을 어떻게 인용하는지를 활용하고, 결과적으로 더 최근의 지식 구조를 분석한다 [39], [40].

한편, 저자서지결합(author bibliographic coupling, ABC)은 두 저자가 그들이 출판한 문헌들에서 같은 저자의 문헌을 인용하는 현상을 이야기한다. 저자서지결합분석은 서지결합의 분석 단위를 문헌 수준에서 저자 수준으로 확장하여, 저자 단위의 서지결합빈도를 활용하여 저자 간 관계를 분석하는 방법이다. 또한,

저자서지결합분석은 저자동시인용분석에서 분석의 단위였던 피인용 저자를 측정 수단으로 삼고, 측정 수단이었던 분석대상 문헌을 작성한 저자를 분석의 단위로 삼는다. 즉, 두 저자가 각각 저술한 문헌에 같은 저자의 문헌을 공통으로 인용하였다면 두 저자의 연구 영역이 유사하다는 가정에 기반을 둔 분석 방법이다 [16]. 이러한 저자서지결합의 기본 가정에 따라, 두 저자가 그들의 문헌 속에 더 많은 공통 인용 문헌이 존재할수록 그들의 연구는 서로 더 관계가 있다고 판단할 수 있다.

서지결합관계의 두 가지 주요한 특징은 결합 강도의 안정성과 문헌 간 정적인 관계이다 [38]. 그러나 분석의 단위를 저자 수준으로 확장하면 이 두 가지 특징은 유지되지 않는다. 문헌 간 결합 강도는 안정적이지 않으며, 분석 대상 저자가 추가로 문헌을 출판함에 따라 변하게 된다. 즉, 문헌 단위 서지결합 결과는 문헌 간 고정적 관계를 반영함과 비교하면, 저자 단위 서지결합 결과는 저자 간 동적인 관계를 반영한다는 점에서 문헌 단위 서지결합과 저자서지결합은 관계의 안정성에 차이가 있다.

서지결합분석의 경우 문헌 간 서지결합빈도가 두 문헌의 출판과 함께 고정되므로, 시간의 흐름에 따른 학문 영역의 변화에 대하여 분석하는 것은 어렵다 [3]. 그러나 저자서지결합분석의 경우, 분석 대상 저자 중 최소 1 인 이상이 추가로 문헌을 출판한다면 이에 따라 서지결합빈도가 계속 변화될 수 있으므로, 문헌 단위 서지결합분석이 가지고 있는 문헌 간 서지결합빈도의 고착 문제를 해결할 수 있다는 장점이 있다. 따라서 저자서지결합분석은 지식 관계의 분석에서 현재의 관계와 시간에 따른 관계의 변화를 연구할 수 있게 하는 방법으로 평가된다 [16], [41].

이재윤은 그의 연구를 통하여, 저자서지결합분석을 통해 나타나는 결과의 전반적인 지적 구조는 저자동시인용분석의 결과와 유사하나, 저자동시인용분석에서 나타나는 인용의 지체 현상이 없으므로 세부적인 측면에서는 저자동시인용분석에 드러나지 않는 최신 연구주제와 현재 활발히 연구를 수행하는 연구자, 새롭게 연구를 시작하고 있는 신진 연구자의 파악이 가능하다고 하였다 [16]. 또한, 우리나라와 같은 인용 수입초과국의 입장에서는 국내 연구자만을 단위로 하는 지적 구조 분석이 가능하므로 국내 동향 분석에 가장 적합한 방식이라는 기법상의 장점이 있다고 하였다.

Zhao 와 Strotmann 은 저자동시인용분석이 최근 20 여 년간 학문 영역의 지적 구조를 표현하는데 가장 많이 사용된 기법이지만, 저자동시인용분석은 학문 영역의 현재와 과거에 대한 지적 구조와 학문 영역의 내부와 외부로 분석하기에 유용한 기법이며 [41], 저자서지결합분석은 학문 영역에서 최신의 활발한 연구 분야를 표현하는 데 있어 매우 적합한 방법이라고 하였다 [17]. 또한, 인용 데이터베이스의 품질 향상으로 문헌의 모든 저자 정보가 포함됨에 따라, 저자서지결합분석이 저자의 지적 구조를 연구함에 제 1 저자에 국한되지 않고 모든 저자를 고려할 수 있게 되었다. 이에 따라 저자서지결합분석을 통하여 지적 구조와 그 특성, 그리고 시간에 따른 관계의 변화에 대한 더 자세한 분석이 가능하게 되었다 [41].

2012 년에 Ma 의 연구에서 활용한 저자서지결합강도의 개념은 문헌서지결합강도에서 유도되었지만, 저자서지결합강도의 실제 계산은 조금 더 복잡하다 [42]. 저자동시인용강도의 개념과 비슷하게, 기초적인 저자서지결합강도는 아래의 3 가지 방법으로 얻을 수 있다. 계산 과정에서 저자의 문헌에 나타난 모든 인용 정보를 고려하지만,

Rousseau 의 방법을 따라 두 저자에 의해 공동 작성된 문헌의 인용 정보는 제외한다 [43].

1. 단순 방법(Simple method): 이 방법은 각각의 저자에게서 동일 문헌이 여러 번 인용되는 것은 고려하지 않고, 공통 인용 문헌의 발생 횟수만을 사용한다. 두 저자 간 저자서지결합의 강도는 두 저자의 인용 목록에 동시에 나타나는 문헌들의 수를 누적하여 계산한다.
2. 최소 방법(Minimum method): 이 방법은 단순 방법과는 달리 각각의 저자에게서 여러 번 인용된 문헌을 고려하며, 각각의 인용 목록에서 나타나는 문헌의 총 인용 횟수는 강도 계산의 가중치로 사용한다. 두 저자 간 저자서지결합의 강도는 두 저자의 인용 목록에 동시에 나타나는 문헌들의 인용 횟수 중 최솟값을 누적하여 계산한다.
3. 혼합 방법(Combined method): 이 방법은 각각의 저자의 인용 목록에서 반복된 인용 문헌이 나타나는 것을 고려하지만, 최소 방법과는 달리 두 저자의 인용 정보에 나타나는 인용 횟수를 서로 곱하여 강도 계산의 가중치로 사용한다. 두 저자 간 저자서지결합의 강도는 위의 가중치를 누적하여 계산한다.

표 1. Ma 의 유사도 도출 방법별 유사 지수 예시

저자	인용 문헌(총 인용 횟수)	공통 인용 문헌
A	문헌 1(1), 문헌 2(1), 문헌 3(3)	문헌 1(A 1 회, B 2 회) 문헌 3(A 3 회, B 2 회)
B	문헌 1(2), 문헌 3(2), 문헌 4(1)	
단순 방법 유사도		최소 방법 유사도
$1 + 1 = 2$		$\min(1,2) + \min(3,2)$ $= 1 + 2 = 3$
		혼합 방법 유사도
		$1 \times 2 + 3 \times 2 = 8$

Ma 는 위의 세 가지 방법을 통하여 도출된 결과를 실험적으로 비교하였다. 비교 결과에 따르면, 단순 방법은 유사성 측정이 가능한 필수 문헌의 반복 인용을 고려하지 않아 두 저자 간 관계를 축소한다. 반면 혼합 방법은 반복 인용된 문헌의 역할을 과대평가하여 두 저자 간 관계를 과장한다. 그러나 최소 방법은 위 두 방법의 중간에 위치하며 반복된 인용 문헌의 역할을 축소하거나 과장하지 않는다. Ma 는 유사도 도출 방법 중 최소 방법이 가장 적절하다고 결론지었다.

2.3.2 선행 연구

서지결합은 문헌 간 연관성 측정 방법으로서 동시인용보다 먼저 제안되었다. 그러나 동시인용분석이 다양한 학문 영역의 지적 구조와 그 변화를 표현하기 위해 주로 사용된 것과는 다르게, 서지결합은 연구 평가나 지적 구조 분석에 적용된 사례가 많지 않다. 이는 인용 분석의 기초 자료가 되는 인용 데이터베이스에서 서지결합정보를 곧바로 얻어내기 어렵거나 불가능했던 것에 일부 원인이 있었다. 그러나 인용 데이터베이스의 품질 향상과 함께 이러한 어려움이 해소됨에 따라, 학문 영역의 지적 구조를 분석하기 위해 서지결합분석을 적용한 연구들이 최근에 다시 나타나고 있다. 이는 서지결합분석이 앞서 언급한 동시인용분석에서 필연적으로 발생하는 몇 가지 문제를 회피할 수 있게 하는 장점이 있기 때문이다.

Glanzel 과 Czerwon 은 서지결합법이 새로운 주제를 파악하는 데 유용하지만 널리 사용되지 않는 것이 놀랍다고 하였다 [44]. Morris 는 서지결합법이 같은 주제를 가지는 문헌 군집을 생성할 때 유용하다고 하였으며 [45], Huang 등은 신속한 연구 동향 파악이 가능하다는

장점을 활용하여 서지결합법으로 특허인용지도를 생성하기도 하였다 [46].

Kuusi 와 Meyer 의 나노튜브와 관련 특허에 관한 서지결합 패턴 사례 연구에서는 서지결합이 기술적 돌파구를 예측하는데 적절하다고 하였다 [47]. 또한, 동시인용분석이 강한 신호에 기반을 두어 활발한 연구 영역을 감지하는데 초점을 둔 것과는 달리, 서지결합분석에서는 연구 대상 문헌을 문헌의 인용 횟수에 의해 선택하지 않기 때문에 약한 신호에 기반을 둔 연구 동향도 감지될 수 있다는 것이 몇몇 연구를 통해 발견되었다 [44], [48].

서지결합방법을 활용한 측정 결과에 기반을 둔 지적 구조 분석은 Jarneving 에 의해 문헌 단위로 [49], [50], Boyack 에 의해 저널 단위로 수행되었다 [51]. 또한, 연구 동향 추출과 과학을 조망하는 데 대한 문헌 서지 결합 사용의 효과는 관련 연구를 통하여 증명되어왔다 [50], [52].

한편, 저자서지결합분석에 대한 연구 역시 저자동시인용분석에 비하여 비교적 적은 편이다.

White 는 8 명의 정보과학자가 그들의 공통적인 관심 연구 영역을 보여주는 많은 수의 같은 저자들을 인용했다는 것을 언급했다 [53]. 이것이 저자서지결합의 개념과 정확히 일치하는 것은 아니지만, 저자를 단위로 하여 저자 간 유사성에 대한 분석을 시도한 측면이 있다. Leydesdorff 는 그의 개인 웹사이트에 저자서지결합을 위한 소프트웨어를 공개하였다 [54]. Zhao 와 Strotmann 은 문헌정보학 영역에 저자서지결합분석을 시도하여, 저자서지결합분석이 저자들을 통해 영역의 지적 구조를 연구하는 데 추가적인 정보를 제공한다고 하였다 [17]. Rousseau 는 저자결합강도를 이론적으로 계산하고, 단순

결합 강도와 상대적 단순 결합 강도의 두 가지 계산 방법을 제안하였다 [43].

Ma 는 Chinese Social Sciences Citation Index(CSSCI)의 문헌정보학 분야를 대상으로 저자서지결합분석을 수행하였다 [42]. 연구를 통해 그는 저자동시인용분석에 비교한 저자서지결합분석의 장점을 보이고, 저자서지결합강도를 계산하는 적절한 방법을 제안하였다. 또한, 분석 결과의 시각화 과정에서 고려할 요소들에 대해 언급하였다.

국내에서는 2006 년 이재윤이 “서지적 저자결합분석”이라는 이름으로 저자서지결합분석을 처음으로 제안하였고, 이후 2008 년에 학술지에 발표하였다 [16]. 그는 영양학 분야와 건축학 분야를 다룬 두 가지 기존 저자동시인용분석 연구에 대하여 저자서지결합분석을 활용하여 그 결과를 재해석하고 효용성을 검증하였다.

김희전과 조현양은 사회복지학 분야의 지적 구조를 분석하는 데 있어 저자동시인용분석과 저자서지결합분석의 조합의 유용성을 분석하였다 [55]. 그들은 1999 년부터 2009 년까지의 사회복지학 영역의 국내 저널 2 개에 수록된 문헌의 인용 정보를 수집하고, 그중 18 번 이상 인용된 46 명의 저자에 대한 분석을 수행하였다. 연구 결과, 저자서지결합분석을 통하여 사회복지학 영역의 지적 구조뿐만 아니라 최신의 연구 동향 및 최근 활발한 연구 활동을 하는 핵심 연구자를 파악할 수 있었다. 그들은 결론에서 특정 학문 영역의 지적 구조를 분석하는 데 있어 저자동시인용분석뿐만 아니라 저자서지결합분석을 병행하는 것이 학문의 하위 주제영역의 파악이나 연구자들의 연구 경향을 파악하는 데 효율적이라고 주장하였다.

이후 저자서지결합분석을 활용하여 특정 학문 분야의 지적 구조를 분석하고자 하는 연구가 점차 증가 추세에 있다. 변지혜와 정은경은

전기공학 분야에 대하여 [56], 김희진은 한국 미술치료학 분야에 대하여 [57], 민형진은 국내 IT 아웃소싱 연구에 대하여 [58], 그리고 박지연과 정동열은 문헌정보학 분야에 대하여 [59] 저자서지결합분석을 통하여 지적 구조를 분석하고 그 유용성을 평가하였다.

2.4 네트워크 분석 기법

계량서지학에서는 다양한 방법으로 학문 또는 산업 영역의 지적 구조를 분석하는 연구를 수행하고 있다. 계량서지적 분석의 조사 대상은 그 규모에 따라서 문헌, 특허, 웹페이지 또는 저자를 연구 단위로 하는 미시적 수준(microlevel), 학교, 연구소, 기업 또는 학술지를 연구 단위로 하는 중간 수준(mesolevel), 그리고 국가 또는 학문 영역 등을 연구 단위로 하는 거시적 수준(macrolevel)으로 분류된다 [60].

계량서지적 분석을 위한 방법으로는 Egghe 가 언급한 양적인 네트워크 분석이 최근 새로운 기법으로 주목되어 그 활용 빈도가 증가하고 있다 [61], [62]. 특히 지적 구조의 시각적인 표현을 통한 분석을 위하여 여러 가지 관계 표현 기법이 도입되고 있다 [63]. 이러한 관계 표현 기법을 통해 표현된 지적 지도상에서 군집은 해당 영역 내에서 세부 주제 영역을 파악할 수 있게 하는 유용한 수단이 된다 [15].

관계 행렬로부터 일부 연결관계를 선택하거나 제거하여 시각적으로 유의미한 정보를 도출할 수 있게 하는 관계 표현 알고리즘은 여러 가지가 있다. 학문 영역이나 연구 영역의 지적 구조를 분석하는 전통적인 연구에서는 다차원척도법과 군집분석을 주로 활용하여

시각화 및 분석을 수행해왔으나, 최근에는 패스파인더 네트워크 알고리즘을 활용하여 다차원척도법 또는 군집분석을 보완하거나 대체하는 연구가 증가하고 있다.

저자서지결합분석을 수행하기 위해서는 먼저 각각의 두 저자의 유사도를 원소로 하는 유사도 행렬을 구하고, 이를 군집분석, 요인분석, 그리고 다차원척도법 등을 사용하여 분석 및 시각화하여 결과를 얻는다 [20]. 그러나 유사도 행렬의 생성 방법에 있어, 특히 적절한 저자 간 유사도 계산 방법에 대한 다양한 의견이 제시되었다. White 등의 연구자들은 전통적으로 활용되는 유사도 지수인 피어슨 상관계수를 지지하였다 [12]. Ahlgren, Jarneving, 그리고 Rousseau 등의 연구자들은 코사인 유사도 등의 다른 지수들을 활용하여 피어슨 상관계수를 대체하고자 하였다 [22]. 또한, White 등의 연구자들은 기존의 빈도행렬에 기반을 둔 다른 데이터 마이닝 기법을 탐색하였다 [23]. White 는 문헌정보학 영역에서 많이 인용된 120 명의 저자에 대해 패스파인더 알고리즘을 활용하여 지적 지도를 생성하였다. 또한, 그는 패스파인더 알고리즘에서 원래의 동시 발생 강도 행렬을 사용하는 것이 피어슨 상관계수를 사용하는 것에 비해 결과가 더 현실적이고 설득력이 있다는 것을 증명하였다.

2.4.1 다차원척도법

다차원척도법(Multi Dimension Scaling, MDS)과 군집분석은 White 와 Griffith 가 저자동시인용분석에서 사용한 후 지적 구조의 시각화 방법으로 널리 사용되어왔다 [10]. 다차원척도법과 같은 차원축소 방법을 통하여 분석 대상의 차원을 2 차원까지 축소하게 되면 다변량 분석의 결과를 바로 시각적으로 표현할 수 있게 된다. 이렇게

다차원척도법과 군집분석은 각각의 문헌 또는 저자의 군집을 표현하면서 동시에 전체적인 지적 구조를 보여줄 수 있는 장점을 가지고 있어 연구에 자주 활용되었다.

그러나 최근에는 다차원척도법이 지적 구조의 시각화 방법으로 한계가 있음이 지적되고 있다 [63]. 주장을 정리하면 다음과 같다.

첫째, 표현해야 할 대상이 너무 많으면 SPSS 와 같은 통계 프로그램의 다차원척도분석 모듈을 활용할 수 없다. 만약 활용할 수 있게 되더라도 개체가 너무 많으면 개체 간 거리의 변형을 뜻하는 스트레스 값이 매우 커지게 되어 왜곡이 심해진다.

둘째, 수십 개 이상의 개체를 2 차원 MDS 지도로 표현하게 되면 세부 구조의 표현력이 매우 저하된다 [64], [65]. 전체적인 개체들의 배치는 적절하게 표현되더라도 각 군집 내에서는 개체 간 거리가 상관관계를 제대로 표현하지 못하는 경우가 많다. 실제 연구 결과에서, MDS 지도상에서 특정 개체와 가장 가깝게 위치한 개체가 상관계수 상으로는 가장 가깝지 않은 경우가 종종 나타났다. 이를 극복하기 위해 Ding 등은 전체 MDS 지도는 군집 단위로 표시하고 각 군집 별로 세부 MDS 지도를 생성하는 다단계 전략을 제안하기도 하였다 [66].

셋째, 표현된 MDS 지도를 직관적으로 해석하기가 어렵다 [67]. 다차원척도법의 결과로 생성되는 MDS 지도에서 가로축과 세로축은 연구자가 개체의 배치 상태 분석을 통하여 의미를 부여하여야 하며, 그 자체로는 의미를 지니지 않는다.

넷째, MDS 지도만으로는 소주제 집단을 식별할 수 없다. 이 때문에 MDS 지도와 군집분석을 함께 활용하는 경우가 많은데, 이 경우에도 MDS 지도가 세부 구조의 표현력이 떨어지므로 실제 가까운 교점과

지도상 가깝게 나타나는 교점이 일치하지 않아서 구조를 파악하기 어려운 군집이 그려지는 경우가 있다.

2.4.2 패스파인더 네트워크

지적 구조를 시각화하는 다른 방법으로는 네트워크 표현 방법이 있다. 네트워크 표현은 분석 단위를 교점(node)으로, 분석 단위 간 관계를 연결(link)로 나타내는 방법으로서, 다차원척도법보다 지적 구조의 표현에 사용되어온 역사가 오래되었다. 문헌동시인용분석을 제안한 Small 이 결과의 시각적 표현을 위하여 네트워크 표현 방법을 활용하였으며 [3], 이후 1970 년대 말까지는 개별 문헌네트워크와 문헌군집네트워크를 별도로 생성하여 지적 구조를 분석하였다 [68].

패스파인더 네트워크 알고리즘(Pathfinder Network Algorithm)은 인지심리학자인 Schvaneveldt 가 1980 년에 발표한 논문과 1990 년에 출판한 단행본에서 제시한 알고리즘이다 [64]. 이 네트워크 표현 방법은 원래는 다차원척도법처럼 심리학 연구에서 유사성 자료를 분석하기 위한 도구로 개발되었으며, 다차원척도법의 한계가 드러나면서 지적 구조의 시각적 표현을 위한 대안으로 다시 주목을 받게 되었다.

패스파인더 네트워크는 최소신장트리(minimum spanning tree, MST)와 삼각부등식(triangle inequality)의 정리에 기반을 두고 있다 [64]. 패스파인더 네트워크를 생성하는 알고리즘을 패스파인더 네트워크 알고리즘이라고 하며, 다변량 분석 기법의 일종으로 간주하여 패스파인더 네트워크 척도법(Pathfinder Network Scaling) 또는 패스파인더 척도법이라고 부르기도 한다 [69].

패스파인더 네트워크는 관계가 설정된 모든 연결이 생성된 상태에서 삼각부등식(triangle inequality)을 위반하는 경로를 제거하여 생성되는 네트워크이다 [64]. 삼각부등식이 위반되는 경우란, 직접 연결되는 긴 연결보다 여러 개의 짧은 연결을 통해 간접적으로 연결되는 경로가 존재하는 경우를 말한다.

패스파인더 네트워크에서의 삼각부등식 위반 여부를 결정하기 위해서는 두 가지 변수 q 와 r 이 필요하다. 변수 q 는 교점 사이의 거리를 산출하는 데 고려하는 최대 연결 수를 의미하며, 2 에서 $n-1$ (n 은 교점의 총 수)까지 설정할 수 있다. $q=2$ 이면 한 교점을 거쳐 우회하는 경로만 고려하면 되고, $q=n-1$ 이면 다른 모든 교점을 거쳐 우회하는 경우까지 모두 고려해야 한다. q 가 커질수록 조사대상 범위가 넓어져서 엄격한 조건이 되어 네트워크에 남는 연결의 수가 줄어든다.

변수 r 은 민코프스키 거리 공식의 제곱 계수로서 한 경로를 구성하는 여러 연결의 가중치를 거리에 반영하는 방법을 결정한다. 민코프스키 거리는 경로를 구성하는 각 연결의 가중치를 r 제곱하여 합한 다음 r 제곱근을 취한 값이다. $r=1$ 이면 각 연결 가중치의 합이 그대로 경로의 거리가 되고, $r=\infty$ 이면 경로를 구성하는 연결의 가중치 중 최댓값이 경로의 거리가 된다. r 이 커질수록 경로의 거리가 짧아지므로 엄격한 조건이 되어 네트워크에 남는 연결의 수가 줄어든다.

패스파인더 네트워크로 지적 구조를 표현하기 위해서는 주로 가장 엄격한 조건인 $r=\infty$, $q=n-1$ 으로 설정(PFNet($r=\infty$, $q=n-1$) 이라고 표현)하여 주요 흐름이 표현되도록 한다. 이때 패스파인더 네트워크는 모든 최소신장트리를 합친 것과 같다 [70]. 즉, 연결 가중치가 동등일 경우 임의의 연결을 선택하는 최소신장트리와는 달리, 동등인 연결을

모두 선택하는 경우가 $\text{PFNet}(r=\infty, q=n-1)$ 에 해당한다. 따라서 최소신장트리와 패스파인더 네트워크는 유사한 점이 많으며, 만약 각 연결의 가중치가 동률인 경우가 없다면 최소신장트리와 $\text{PFNet}(r=\infty, q=n-1)$ 는 같아진다.

따라서 최소신장트리와 패스파인더 네트워크는 개별 교점의 전체적인 배치 및 구조 측면에서 비슷한 결과를 보이며, 결과 네트워크에서의 전체적인 흐름에는 큰 차이가 없다. 그러나 세부 구조 측면에서는 패스파인더 네트워크 알고리즘을 사용한 경우가 최소신장트리를 사용한 경우에 비해 세부 구조가 더 잘 나타나는 경우가 있다. 이는 최소신장트리의 경우 동률인 연결 중 임의로 하나를 선택하므로 세부 연결의 빠지는 부분이 발생하기 때문이다.

계량서지학적 분석에 패스파인더 네트워크 알고리즘을 적용하였던 연구자들은 패스파인더 네트워크 알고리즘이 다차원척도법보다 세부 구조가 잘 나타날 뿐만 아니라 전체적인 구조도 더 뚜렷하게 제시해준다고 평가하였다 [23], [71], [72]. 또한, 전체 교점이 하나의 네트워크를 형성하는 것이 보장되므로 분석 대상 전체에 대한 관계 지도를 생성할 수 있다. 이에 따라, 다차원척도법을 보완하거나 대체하는 용도로 패스파인더 네트워크 알고리즘을 사용하는 경우가 증가하고 있다.

그러나 패스파인더 네트워크는 최소신장트리에 비해 알고리즘의 복잡도가 높아 추가적인 처리시간과 기억공간이 필요하다 [73]. Schvaneveldt 가 제안한 알고리즘의 계산복잡도는 $O(N^4)$ 로 매우 높는데, 이는 행렬 전체에 대해 삼각부등식을 위반하는지를 검사해야 하기 때문이다 [70]. 특히 지적 구조 분석 연구에서는 주요 연결만을 남기기 위해 가장 연산량이 많은 $\text{PFNet}(r=\infty, q=n-1)$ 을 주로 적용하기

때문에, 분석 대상 교점이 수백 개 이상이면 처리 시간이 매우 많이 필요하게 된다 [73].

또한, 패스파인더 네트워크상에서 군집을 형성하여 주제를 파악할 방법이 없는 것은 패스파인더 네트워크 알고리즘의 활용을 제한하고 있다. 이 때문에 패스파인더 네트워크 알고리즘을 적용한 연구에서도 다차원 척도법에 따라 MDS 지도상에 군집 분석 결과를 표시하는 전통적인 방식을 병행하는 경우가 있다 [74], [75]. Chen 은 패스파인더 네트워크 표현과 군집분석이 잘 결합하지 않는 것을 해결해야 할 과제 중 하나로 언급하고 있다 [70].

McCain 은 생물공학 분야의 특허에 관한 연구를 수행하면서 특허 항목의 동시분류 분석에 [76], Chen, White 등은 저자동시인용분석에 [23], [77], Marion 과 McCain 은 소프트웨어공학 분야 저널동시인용분석에 패스파인더 네트워크 알고리즘을 도입하였다 [74]. 특히 1998 년에 McCain 과 함께 연구를 수행했던 White 가 2003 년에 같은 연구자료에 대해 패스파인더 네트워크 알고리즘을 적용하여 더욱 개선된 연구 결과를 발표한 것은 패스파인더 네트워크 알고리즘의 장점이 널리 알려지는 계기가 되었다 [23].

한편, 1990 년대 후반 Chen 이 패스파인더 네트워크 알고리즘 관련 연구를 발표하고, 계량서지적 자료를 패스파인더 네트워크로 분석할 수 있는 공개 소프트웨어인 CiteSpace 를 개발하면서 패스파인더 네트워크 알고리즘의 보급에 이바지하였다 [78]. Chen 은 패스파인더 네트워크를 그대로 제시하는 것에 그치지 않고 각 교점의 역할을 전환점 역할 교점, 피벗 교점, 그리고 허브 교점 등으로 구분하여 파악할 것을 제안하였다.

2.5 과학지도학

2.5.1 정의 및 특징

과학지도학(Scientography)은 학술 문헌을 활용하여 학문 영역 간 관계를 시각적으로 표현하고 이를 분석하는 일련의 학문을 말한다. 과학지도학은 1966 년 Price 에 의해 최초로 그 개념이 제시되었으며 [79], 1985 년 Small 과 Garfield 의 문헌에서 과학 지도(map of science)가 처음으로 생성되었다 [80]. “과학지도학”이라는 명칭은 1994 년 Garfield 의 문헌에서 처음 명명되었으며 [81], Vargas-Quesada 와 Moya-Anegón 은 학술 활동의 산출물을 활용하여 도식화하는 활동 및 그 효과에 대한 표현으로 “과학지도학”이 “도메인 시각화”나 “정보/지식 시각화”보다 적절한 표현이라고 평가하였다 [82].

Small 은 공간상에 학술 문헌을 배치하는 것이 학문 분야 간 개념 관계와 학문의 전개 과정을 이해하는 데 도움이 되며, 지식의 현황에 대한 통찰을 제공해줄 수 있다고 하였다 [83].

2.5.2 선행 연구

1985 년 Small 과 Garfield 는 1983 년의 SCI 및 SSCI 문헌 50,994 건을 대상으로 과학 지도를 처음으로 생성하였다 [84]. 이 과학 지도는 사회과학 및 자연과학 영역의 문헌을 동시인용 방법을 활용하여 다단계 군집화 또는 다차원척도법(MDS)을 적용하여 시각화하였으며, 광역 수준의 지도와 개별 논문 수준의 지도를 모두 포함하고 있다.

1999 년 Small 은 1995 년의 SCI 및 SSCI 문헌 164,612 건을 대상으로 과학 지도를 생성하였다 [83]. 이 과학 지도 역시 사회과학 및 자연과학 영역의 문헌에 대해 동시인용 방법을 활용하였으며, 다단계 군집화 또는 삼각측량법을 이용하여 시각화하였다.

이후 인용 정보를 제공하는 학술 데이터베이스가 증가함에 따라 과학 지도 생성을 위한 여러 시도가 나타나기 시작했다. Moya-Anegón 등은 2000 년의 Web of Science 의 문헌을 대상으로, 저자 정보의 주소에 스페인이 들어있는 문헌 26,062 건에 대해 과학 지도를 생성하고, 이를 “scientogram”이라고 이름 지었다 [85]. 그들은 학술지간 동시인용 빈도를 산출한 후 각 학문 영역 단위로 합산하여 학문 영역 간 관계를 산출하는 범주 동시인용 방법을 활용하였으며, 최소신장트리를 활용하여 이를 시각화하였다.

2005 년 Boyack 등은 2000 년의 SCI 및 SSCI 학술지 7,121 종을 대상으로 학술지 동시인용 방법을 활용하여 과학 지도를 생성하였다 [86]. 그들은 학술지 간 관계 산출 과정에서 총 8 가지 방법을 적용하여 관계성을 산출하였으며, 그 중 자카드 유사 계수를 활용한 방법이 다른 방법에 비해 근소하게 우수하다고 평가하였다. 도출된 결과는 기존의 ISI 분류표와 비교하여 얼마나 유사한지에 대해 평가하였다.

2007 년 Vargas-Quesada 와 Moya-Anegón 은 2002 년의 Web of Science 의 전 세계 문헌 901,493 건과 스페인 문헌 294,778 건에 대하여 과학 지도를 생성하였다 [82]. 그들은 관계 분석 방법으로 기존에 활용하였던 범주 동시인용 방법을 활용하였으나, 시각화 방법으로는 패스파인더 네트워크를 활용하였다. 또한, 그들은

동시인용빈도가 동물인 경우의 처리를 위해 Salton 지수를 추가하여 활용하였다.

국내 연구의 경우, 개별 학문 영역에 대한 과학 지도를 생성한 연구는 다수 있으나 [26], [35], [55], 학문 전 영역을 대상으로 하여 과학 지도를 생성하고자 하는 시도는 많지 않다. 이재운은 2001 년부터 2005 년까지 한국학술진흥재단에 연구비 지원을 신청한 연구과제 80,502 건을 분석 대상으로 하여 광역 과학 지도 생성을 시도하였다 [87]. 그는 관계 분석 방법으로 자동 추출 색인어의 유사도를 활용하였으며, 시각화 방법으로는 패스파인더 네트워크와 자체 개발한 클러스터링 기반 네트워크를 활용하였다.

국내의 광역 과학 지도 생성 연구가 미비한 원인으로서는 그동안의 인용 데이터베이스 구축의 미비가 지적되었다. KCI 데이터베이스가 2001 년부터 준비를 시작하여 2005 년 구축되었으나, 데이터 수록 기간이 짧고 범위가 제한적이며, 인용 정보가 불완전하고, 인명 정보가 통제되지 않는 등의 문제점이 있었다 [88]. 그러나 이후 추가 데이터 축적 및 데이터 정리 과정을 통하여 이러한 문제들이 점차 해소되고 있다.

제 3 장 연구 문제

선행 연구 조사를 통해, 국내 문헌 또는 인용 데이터베이스를 대상으로 학문 영역 간 관계를 도출하려 시도한 연구는 많지 않으며, 특히 서지결합분석 방법을 활용한 연구는 동시인용분석 방법을 활용한 연구에 비하여 상당히 드물게 이루어진 것을 확인할 수 있었다. 서지결합분석 방법을 활용한 경우, 대부분 연구자에 의해 선택된 특정한 학문 영역 내에서 일정한 제한 기준을 통해 선별된 저자와 그들의 인용 정보를 분석에 활용한 연구였다. 그러나 학문 영역 내에 속하는 문헌 전체를 분석에 활용한 연구는 찾기 어려웠다.

따라서 본 연구에서는 이러한 한계점을 극복하고자 다음과 같은 연구 문제들을 확인하고자 한다.

연구 문제 1. 국내 인용 데이터베이스를 기반으로, 분석 단위를 학문 영역으로 확장하여 서지결합분석을 수행하여 학문 영역 간 유사 관계를 도출하고, 광역 과학 지도를 생성하여 지적 구조를 표현하는 것이 가능한가?

연구 가설 1. 국내 인용 데이터베이스를 활용하여 분석 단위를 학문 영역으로 확장한 서지결합분석 역시 저자 또는 개별 문헌 단위에서와 마찬가지로 학문 영역 간 유사 관계를 도출할 수 있을 것이며, 광역 과학 지도 생성을 통한 지적 구조 표현이 가능할 것이다.

연구 문제 2. 연구 문제 1의 유사 관계 도출의 방법에 있어, 빈도 기반 유사 지수 또는 유사도 기반 유사 지수 중 어떠한 것이 학문 영역 간 지적 구조의 표현에 더 적합한가?

연구 가설 2-1. 유사 지수 도출 방법의 차이에 따르는 학문 영역 간 지적 구조 표현의 차이가 존재할 것이다.

연구 가설 2-2. 도출된 유사 지수 중 학문 영역 간 지적 구조 표현에 더 적합한 지수가 존재할 것이다.

제 4 장 연구 방법

앞에서 설정한 연구 문제들에 대하여 다음과 같은 방법으로 연구를 수행하였다.

4.1 연구 대상 선정

연구를 수행하기 위해서는 국내에서 발행된 문헌들에 대한 참고문헌 정보의 처리가 필수적이다. 또한, 학문 영역 간 관계의 도출을 위해서는 다양한 학문 영역을 포괄하는 분석 기초 데이터가 필요하다.

KCI 데이터베이스의 경우 국내에서 발행되는 주요 학술지들을 학문 영역의 구분 없이 포괄하고 있으며, 2013 년 10 월 기준 학회지 총 1,827 종, 대학부설연구소 발행지 300 종을 포함하여 총 2,127 종의 학술지에 수록된 문헌들의 정보와 그 참고문헌 정보를 제공하고 있다. 이러한 특징이 연구의 대상으로 적합하다고 판단되어 KCI 데이터베이스를 분석 대상으로 선정하였다.

4.2 데이터 수집

분석을 위해 한국연구재단의 협조를 통하여 KCI 등재 문헌 정보를 수집하였다. 수집 정보는 문헌의 제목, 저자, 분류, 등록 연월, 참고문헌의 제목, 저자, 발표 연월이며, 최종적으로 2013 년 10 월 기준 KCI 총 등재 문헌 935,780 건 중 약 88%에 해당하는 825,189 건의 문헌 정보 및 인용 문헌 정보를 수집하였다. 등재 문헌은 대분류 8 개 항목, 중분류 151 개 항목으로 분류되어 있었으며, 등록된 저자는 총 843,855 명, 저자의 문헌 저술관계는 총 2,255,506 건으로 저자 당 평균 약 2.73 건의 KCI 등재 문헌을 저술한 것으로 집계되었다. 또한,

문헌 단위 인용 관계는 총 12,701,517 건으로 문헌 당 평균 약 15.39 건의 인용이 발생한 것으로 집계되었다.

4.3 데이터 사전 처리

수집된 문헌 및 인용 정보의 사전 처리를 통하여 문헌 간 인용 관계를 확인할 수 있도록 전처리 과정을 수행하였다. 수집된 문헌의 구조적 동일성을 활용하여 문헌의 제목, 저자, 분류, 저술연도, 참고문헌 정보 등을 추출하고, 추출된 정보를 관계형 데이터베이스 시스템(relational database management system: RDBMS)에 저장하여 추후 인출이 쉽도록 하였다.

참고문헌 목록의 문헌이 KCI 에 등재된 문헌인 경우, 문헌의 KCI 고유 코드를 획득할 수 있기 때문에 이를 활용하여 참고문헌의 동일성 여부를 확인할 수 있다. 반면, 대다수를 차지하고 있는 KCI에 등재되어 있지 않은 참고문헌의 경우, 문헌의 동일성을 확인할 수 있는 고유 코드가 존재하지 않는다. 그러나 임의의 두 문헌이 같은 문헌으로 판단되는 경우 하나의 고유 코드만을 부여하여 문헌 간 연관성이 빠지지 않도록 하여야 한다. 따라서 문헌의 동일성을 판단할 수 있는 식별자를 일정한 규칙으로 생성하고, 이를 활용하여 중복된 참고문헌을 판단하는 과정이 필요하였다.

참고문헌의 동일성 판단을 위한 식별자로는 SHA1 알고리즘을 이용하여 해시 코드(hash code)를 생성하였다. 해시 코드 생성을 위한 초깃값으로는 참고문헌의 출판연도, 저자, 제목을 사용하여 고정된 길이를 가지는 해시 코드를 생성한 후, 생성된 코드와 같은 코드가 참고문헌 해시 코드 목록에 존재하는 경우 같은 문헌으로 판단하여 추가 코드를 부여하지 않고 이미 부여된 코드에 대응하였다. 이러한

방법을 통하여, 총 11,753,050 건의 참고문헌 중 같다고 판단된 문헌 총 3,214,063 건을 제거한 참고문헌 8,538,987 건을 이후의 서지결합분석에 활용하였다. 최종적으로, KCI 등재 문헌과 비등재 문헌을 모두 포함한 총 9,364,176 건의 문헌이 분석의 기초 정보로써 활용되었다.

4.4 서지결합분석 수행

위 과정을 통해 처리된 데이터를 바탕으로 하여 서지결합분석을 수행하였다. 분석의 단위는 KCI 중분류 영역 총 151 개를 대상으로 하며, 각 영역에 포함된 문헌들의 참고문헌 정보를 바탕으로 각 학문 영역의 프로파일 벡터를 생성하였다. 프로파일 벡터는 각 영역에서 어떠한 참고문헌을 몇 회 인용하였는지에 대한 정보를 나타내며, 각각의 영역에 속해있는 문헌들에 기록된 모든 참고문헌을 대상으로 하여 생성하였다.

각 영역에 대한 프로파일 벡터를 생성한 후, 이를 이용하여 영역 간 비교과정을 통하여 각 영역에 대한 다른 영역들과의 유사 지수를 측정하고 그 결과를 행렬의 형태로 도출하였다. 영역 간 유사 지수 측정 방법으로서, 빈도 행렬 생성을 위한 방법으로는 Ma 가 수행했던 단순 방법(simple method)과 최소 방법(minimum method), 혼합 방법(combined method)을 통하여 관계를 측정하였고, 유사도 행렬 생성을 위한 방법으로는 자카드 유사 계수(Jaccard similarity coefficient)와 코사인 유사도(cosine similarity), 피어슨 상관계수(Pearson correlation coefficient)를 통하여 관계를 측정하였다.

유사 지수 행렬 생성을 위한 영역 간 유사 지수 도출 방법은 아래와 같이 정리되었다.

1. 단순 방법(simple method): 참고문헌의 동시 발생 횟수만을 누적 합하여 산출
2. 최소 방법(minimum method): 참고문헌의 동시 발생 시, 해당 참고문헌에 대한 각 영역의 인용 횟수 중 적은 쪽을 누적 합하여 산출
3. 혼합 방법(combined method): 참고문헌의 동시 발생 시, 해당 참고문헌에 대한 각 영역의 인용 횟수를 서로 곱한 후 누적 합하여 산출
4. 자카드 유사 계수(Jaccard similarity coefficient): 각 영역의 프로파일 벡터의 0이 아닌 성분에 대하여, 다른 영역과의 전체 성분 대비 공통 성분의 비율로 도출. 두 영역 X, Y 에 대하여,

M_{ij} : X 영역의 성분이 i 이고 Y 영역의 성분이 j 인 속성의 수,

(i, j 는 0(존재하지 않는 경우)또는 1(존재하는 경우))

라고 하면, 자카드 유사 계수는 아래 수식으로 도출

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$$

5. 코사인 유사도(cosine similarity): 각 영역의 프로파일 벡터에 대하여 다른 영역과의 코사인 유사도를 도출

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

6. 피어슨 상관계수(Pearson correlation coefficient): 단순 방법을 통해 도출한 빈도 행렬을 기준으로 표본 피어슨 상관계수 도출

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

(단, 유의수준 $\alpha < 0.01$ 인 데이터만 활용)

4.5 관계 시각화

위에서 도출된 관계 행렬들에 대하여 패스파인더 네트워크 알고리즘 $\text{PFNet}(r=\infty, q=n-1)$ 의 적용을 통해 가장 강한 연결 관계만을 표현하도록 하여 영역 간 관계를 시각화하였다. 시각화에는 D3.js library 를 활용하여 관계 지도를 동적으로 생성하였으며, 시각화 결과를 웹 브라우저에서 바로 접근할 수 있도록 하였다.

관계 지수 선택에 따른 차이를 알아보기 위한 시각화 방법으로는 힘 기반 배치 방법(force-directed layout)을 활용하였다. 각 학문 영역을 교점에, 영역 간 관계를 연결에 대응하여 영역 간 관계를 직관적으로 알아볼 수 있도록 표현하였다.

제 5 장 결과 및 논의

5.1 빈도 지수 기반 관계 분석

5.1.1 단순 방법 활용 결과

단순 방법을 활용하여 도출된 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 2. 단순 방법을 활용한 학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	교육학 (사회과학)	생활과학 (자연과학)	25726
2	경제학 (사회과학)	경영학 (사회과학)	21677
3	관광학 (사회과학)	경영학 (사회과학)	21459
4	기타사회과학 (사회과학)	교육학 (사회과학)	19676
5	심리과학 (사회과학)	교육학 (사회과학)	18813
6	교육학 (사회과학)	체육 (예술체육)	18541
7	사회복지학 (사회과학)	교육학 (사회과학)	18526
8	한국어와문학 (인문학)	기타인문학 (인문학)	15391
9	경영학 (사회과학)	교육학 (사회과학)	15379
10	무역학 (사회과학)	경영학 (사회과학)	14809

분석 결과, 대체로 같은 영역 내의 학문 분야 간 관계가 상위 항목에 나타남을 확인할 수 있다. 특히 사회과학 또는 인문학 영역에 속하는 학문 분야들이 상위 항목의 다수를 차지하고 있음을 알 수 있다.

한편, 서로 다른 학문 영역 간 관계를 살펴보기 위해서는 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수를 검토해볼 수 있다. 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 3. 단순 방법을 활용한 상위 대분류 영역이 서로 다른

학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	교육학 (사회과학)	생활과학 (자연과학)	25726
6	교육학 (사회과학)	체육 (예술체육)	18541
11	교육학 (사회과학)	한국어와문학 (인문학)	13431
13	기타사회과학 (사회과학)	생활과학 (자연과학)	12433
17	사회복지학 (사회과학)	생활과학 (자연과학)	11763
18	관광학 (사회과학)	체육 (예술체육)	11738
19	경영학 (사회과학)	체육 (예술체육)	11144
20	식품과학 (농수해양)	생활과학 (자연과학)	10996
22	심리과학 (사회과학)	생활과학 (자연과학)	9797
24	경영학 (사회과학)	생활과학 (자연과학)	9512

분석 결과, 사회과학 영역에서는 교육학 분야, 자연과학 영역에서는 생활과학 분야, 농수해양 영역에서는 식품과학 분야, 예술체육 영역에서는 체육 분야가 주로 상위 항목에 나타남을 확인할 수 있다. 추가로, 공학 영역의 경우 산업공학 분야에서 경영학 분야와의 관계가 처음으로 나타났으며(유사 지수 6741, 45 위), 의약학 영역의 경우 정신과학 분야에서 교육학 분야와의 관계가 처음으로 나타나는 등(유사 지수 4183, 88 위) 상대적으로 낮은 유사도를 보였다. 이는 국내 학술 환경의 특성상, 공학 영역 및 의약학 영역 연구자들의 국내 발표된 문헌의 수가 많지 않아 분석 대상 문헌 수가 상대적으로 적고, 이는 타 영역과의 참고문헌의 중복 빈도가 확률적으로 낮아지게 되어 유사도가 낮게 나타나는 것으로 추정할 수 있다. 또한, 분야 내 전문성이 강하여

다른 학문 분야에서 문헌을 인용하는 것이 어려운 점도 유사도가 낮게 나타나는 원인으로 추정할 수 있다.

단순 방법을 통해 도출된 학문 분야 간 관계를 시각화한 결과는 다음과 같다.

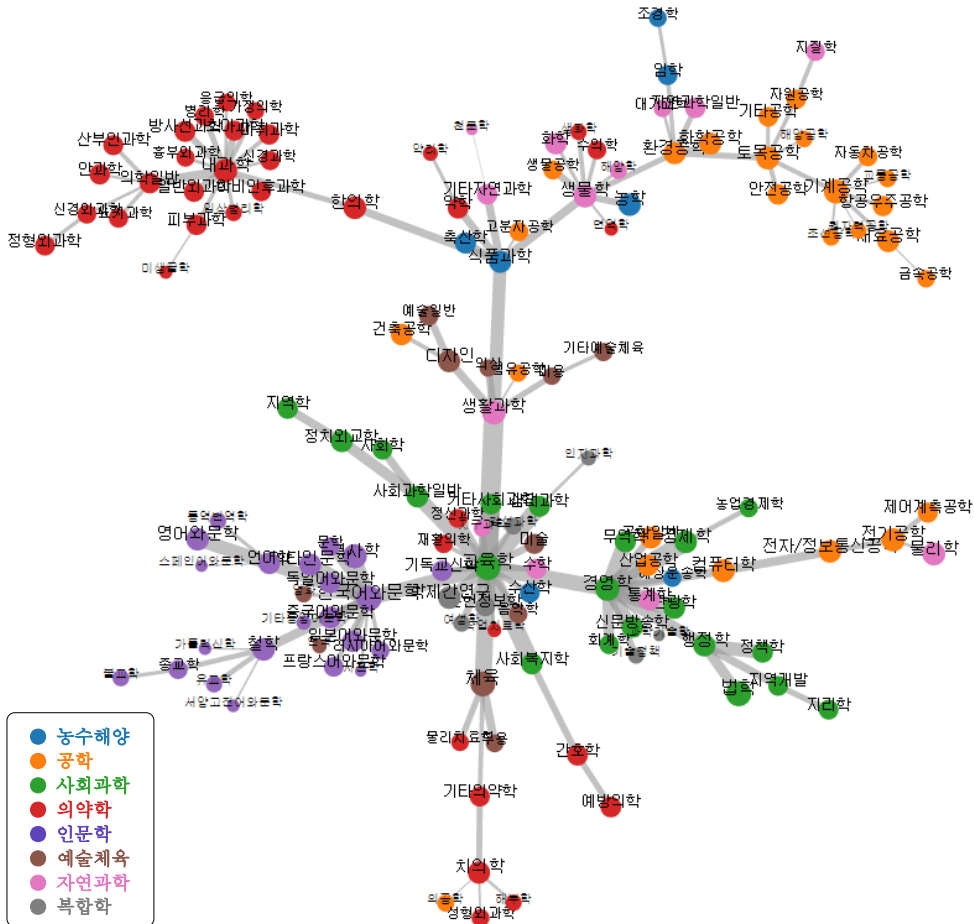


그림 1. 단순 방법을 활용한 학문 분야 간 패스파인더 네트워크

교점의 색상은 각 학문 분야의 대분류 영역에 따라 농수해양(파란색), 공학(주황색), 사회과학(초록색), 의약학(빨간색), 인문학(보라색), 예술체육(갈색), 자연과학(분홍색), 복합학(회색)으로 구분되었다. 시각화 결과 그래프의 중심에 있는 교육학 분야 및 생활과학 분야를

기준으로 공학/의약학 영역과 인문학/사회과학 영역이 대체로 분리되어 나타난 것을 확인할 수 있다.

5.1.2 최소 방법 활용 결과

최소 방법을 활용한 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 4. 최소 방법을 활용한 학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	관광학 (사회과학)	경영학 (사회과학)	38706
2	교육학 (사회과학)	생활과학 (자연과학)	35513
3	경제학 (사회과학)	경영학 (사회과학)	28820
4	교육학 (사회과학)	체육 (예술체육)	27185
5	심리과학 (사회과학)	교육학 (사회과학)	26728
6	기타사회과학 (사회과학)	교육학 (사회과학)	26082
7	사회복지학 (사회과학)	교육학 (사회과학)	24620
8	경영학 (사회과학)	교육학 (사회과학)	22225
9	관광학 (사회과학)	체육 (예술체육)	21070
10	경영학 (사회과학)	체육 (예술체육)	21014

분석 결과, 단순 방법 활용 시와 마찬가지로 같은 대분류 영역 내의 학문 분야 간 관계가 상위 항목에 나타남을 확인할 수 있다. 사회과학 영역에 속하는 학문 분야들이 상위 항목의 다수를 차지하고 있음은 단순 방법 활용 시와 같지만, 예술체육 영역의 체육 분야가 상위 항목에 좀 더 많이 나타나는 차이점이 있다.

한편, 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 5. 최소 방법을 활용한 상위 대분류 영역이 서로 다른

학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
2	교육학 (사회과학)	생활과학 (자연과학)	35513
4	교육학 (사회과학)	체육 (예술체육)	27185
9	관광학 (사회과학)	체육 (예술체육)	21070
10	경영학 (사회과학)	체육 (예술체육)	21014
14	교육학 (사회과학)	한국어와문학 (인문학)	18236
16	기타사회과학 (사회과학)	생활과학 (자연과학)	16559
17	사회복지학 (사회과학)	생활과학 (자연과학)	15910
20	식품과학 (농수해양)	생활과학 (자연과학)	15137
24	경영학 (사회과학)	생활과학 (자연과학)	14037
25	체육 (예술체육)	생활과학 (자연과학)	13504

분석 결과, 단순 방법 활용 시와 마찬가지로 사회과학 영역에서는 교육학 분야, 자연과학 영역에서는 생활과학 분야, 농수해양 영역에서는 식품과학 분야, 예술체육 영역에서는 체육 분야가 주로 상위 항목에 나타남을 확인할 수 있다. 공학 영역과 의약학 영역의 경우 여전히 다른 학문 분야와의 유사 지수가 상대적으로 낮게 나타났다.

최소 방법을 활용한 학문 분야 간 관계의 시각화 결과는 다음과 같다.

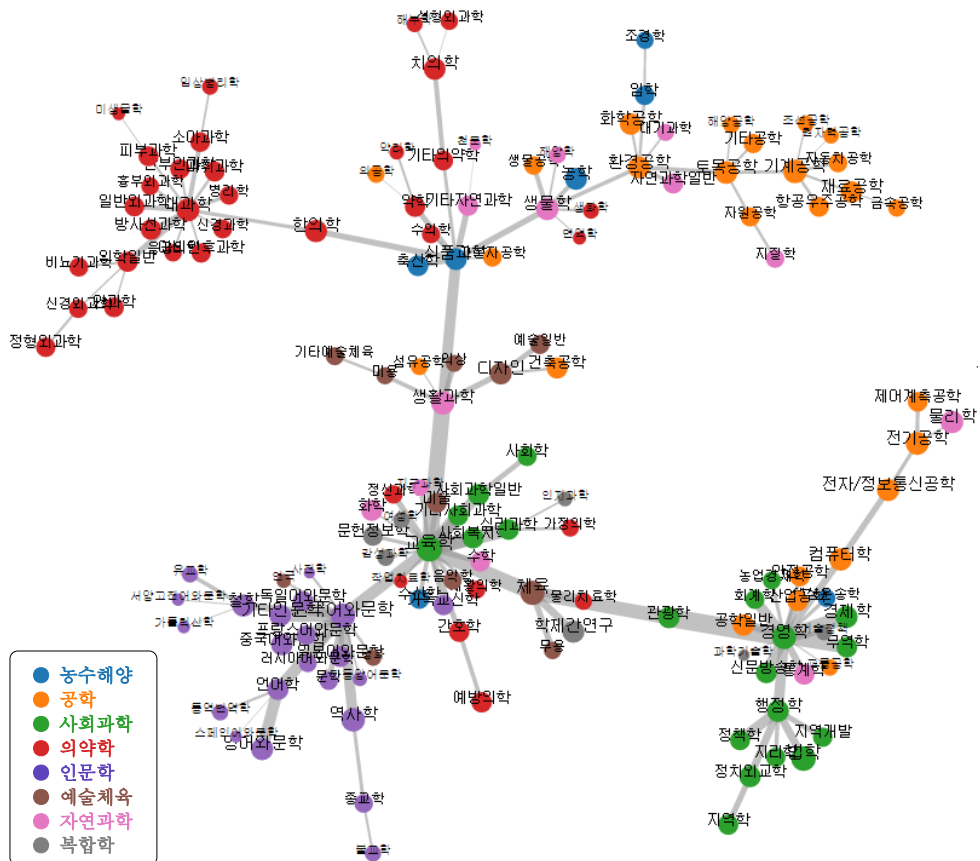


그림 2. 최소 방법을 활용한 학문 분야 간 패스파인더 네트워크

시각화 결과, 전체적인 형태는 단순 방법 활용 시와 비슷하게 나타났다. 그래프의 중심에 있는 교육학 및 생활과학, 식품과학 분야를 기준으로 공학/의약학 영역과 인문학/사회과학 영역이 대체로 분리되어 나타난 것을 확인할 수 있다.

5.1.3 혼합 방법 활용 결과

혼합 방법을 활용한 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 6. 혼합 방법을 활용한 학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	관광학 (사회과학)	경영학 (사회과학)	543077
2	경영학 (사회과학)	체육 (예술체육)	360588
3	교육학 (사회과학)	체육 (예술체육)	296638
4	관광학 (사회과학)	체육 (예술체육)	263036
5	경영학 (사회과학)	교육학 (사회과학)	259167
6	무역학 (사회과학)	경영학 (사회과학)	215540
7	경제학 (사회과학)	경영학 (사회과학)	211775
8	교육학 (사회과학)	생활과학 (자연과학)	207886
9	심리과학 (사회과학)	교육학 (사회과학)	192339
10	한국어와문학 (인문학)	역사학 (인문학)	174105

분석 결과, 유사 지수의 절대적인 크기가 앞서 확인한 단순 방법 또는 최소 방법보다 매우 크게 나타났다. 혼합 방법 활용 시에는 앞의 두 방법과는 달리 사회과학 영역의 경영학, 교육학, 관광학 분야와 예술체육 영역의 체육 분야의 연관성이 상위 항목에 나타나는 차이점이 있다.

한편, 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 7. 혼합 방법을 활용한 상위 대분류 영역이 서로 다른

학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
2	경영학 (사회과학)	체육 (예술체육)	360588
3	교육학 (사회과학)	체육 (예술체육)	296638
4	관광학 (사회과학)	체육 (예술체육)	263036
8	교육학 (사회과학)	생활과학 (자연과학)	207886
14	경영학 (사회과학)	생활과학 (자연과학)	155331
21	교육학 (사회과학)	한국어와문학 (인문학)	127639
22	심리과학 (사회과학)	체육 (예술체육)	102767
23	체육 (예술체육)	생활과학 (자연과학)	98851
24	관광학 (사회과학)	생활과학 (자연과학)	97241
29	교육학(사회과학)	영어와문학(인문학)	76096

분석 결과, 사회과학 영역에서는 경영학/교육학/관광학 분야, 자연과학 영역에서는 생활과학 분야, 예술체육 영역에서는 체육 분야가 상위 항목에 나타남을 확인할 수 있다. 공학 영역과 의약학 영역의 경우 여전히 다른 학문 분야와의 유사 지수가 상대적으로 낮게 나타났다.

혼합 방법을 활용한 학문 분야 간 관계의 시각화 결과는 다음과 같다.

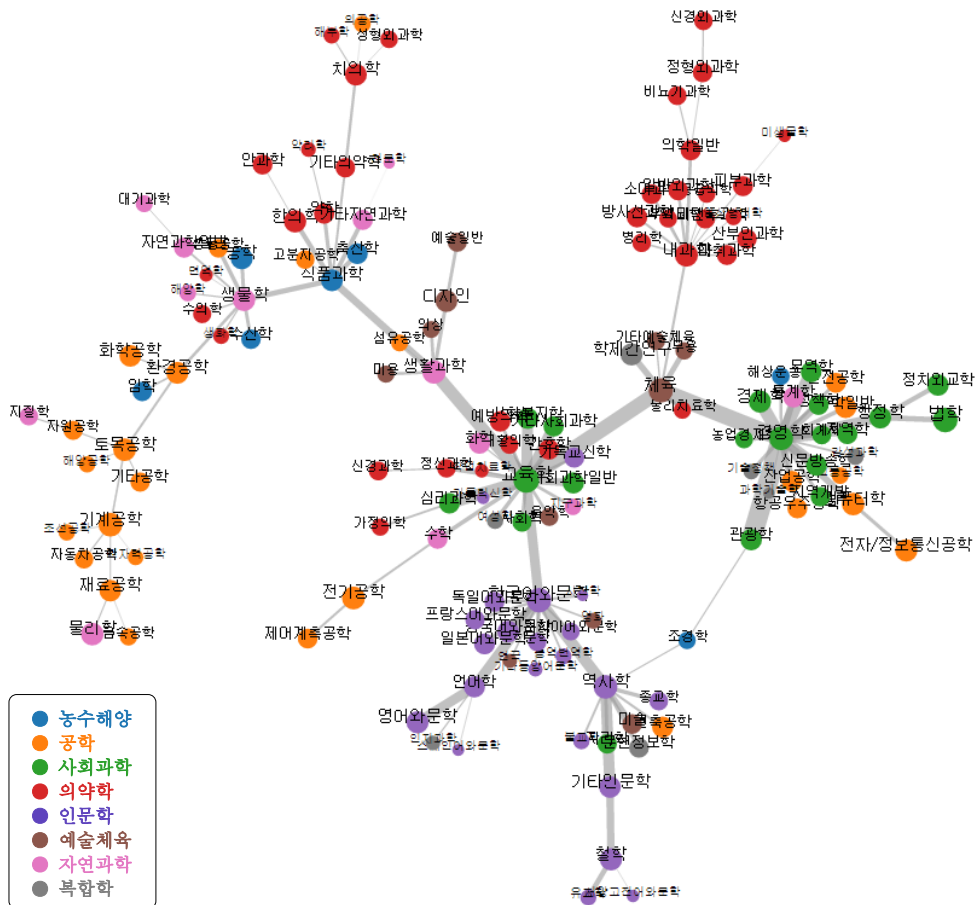


그림 3. 혼합 방법을 활용한 학문 분야 간 패스파인더 네트워크

시각화 결과, 이전의 단순 방법 또는 최소 방법을 활용한 시각화 결과와는 다소 다른 형태로 나타났다. 전체적으로 각 군집에서의 허브 역할을 하는 주된 학문 분야들은 같지만, 다른 분야들과의 세부적인 연결 관계에서는 상당한 차이를 보인다.

5.1.4 빈도 지수 기반 결과의 상호 비교

위의 세 가지 방법으로 도출한 빈도 지수 기반 결과들의 공통된 특징은 다음과 같이 나타났다.

1. 세 가지 방법 모두 학문 분야 간 관계를 시각화하는 것이 가능하였으며, 군집 별로 중심이 되는 학문 분야를 직관적으로 판단할 수 있었다.
2. KCI 대분류 영역 중 인문학 영역과 의학 영역의 세부 학문 분야들은 동일 영역에 속한 학문 분야들과의 유사 관계가 다른 영역에 비하여 세 가지 방법에서 모두 상당히 높게 나타났다.
3. 교육학 분야는 세 가지 방법 모두에서 다른 세부 학문 분야와의 연결도(degree)가 가장 높게 나타났다. 또한, 자신이 속하지 않은 다른 대분류 영역의 세부 학문 분야와의 연결도도 가장 높게 나타났다.
4. 공학 영역은 크게 두 개의 군집으로 분리되어 나타났다. 첫 번째 군집은 전자정보통신공학/전기공학/제어계측공학/컴퓨터학 분야들의 군집으로, 다른 군집과 가장 근접한 학문 분야는 경영학으로 나타났다. 두 번째 군집은 기계공학/토목공학/환경공학 관련 분야들의 군집으로, 다른 군집과 가장 근접한 학문 분야는 생물학으로 나타났다.

그러나 선택한 방법에 따른 결과의 차이점 또한 존재하였다.

1. 단순 방법과 최소 방법의 경우 시각화 결과의 전체적인 구조는 서로 비슷하게 나타났으나, 세부 구조는 몇몇 학문 분야들의 위치와 관계 측면에서 약간의 차이를 보였다.
2. 혼합 방법의 경우, 다른 두 방법에 비하여 학문 분야 간 관계가 직관적으로 해석되지 않는 경우가 발견되었으며, 동일 대분류 영역에 속한 학문 분야들이 다른 방법에 비해 여러 군집에 분산되어 나타나는 경우가 발견되었다. 이는 혼합 방법의 특성에서 기인한 것으로, 혼합 방법의 경우 서지결합빈도 계산

과정에서 공통 인용 문헌의 각 학문 분야별 인용 횟수를 서로
곱하게 되는데, 이 때문에 학문 분야에서 반복적으로 인용된
참고문헌의 영향력이 다른 방법에 비해 과장되는 경향이 있다.
따라서 소수 문헌의 빈번한 공통 인용이 유사 관계에 미치는
영향이 강하게 나타나게 되며, 이는 연결 구조의 차이가 발생하는
원인이 된다.

위의 각 방법에 따른 도출 결과의 공통점과 차이점을 바탕으로, 빈도
지수를 통하여 학문 영역 간 관계를 분석하여 시각화하는 경우 단순
방법 또는 최소 방법을 활용하는 것이 혼합 방법을 활용하는 것에
비해 구조를 분석하는 데 유리하다고 평가할 수 있다. 이는 특정
영역에서의 위의 각 방법으로 도출된 지수를 활용한 시각화 결과에
대하여 해석하였던 Ma의 연구와도 합치하는 결과이다.

5.2 유사도 지수 기반 관계 분석

5.2.1 자카드 유사 계수 활용 결과

자카드 유사 계수를 활용한 학문 분야 간 유사 지수 상위 10 항목은
다음과 같다.

표 8. 자카드 유사 계수를 활용한 학문 분야 간 유사 지수
(상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	영어와문학 (인문학)	언어학 (인문학)	0.0540
2	정책학 (사회과학)	행정학 (사회과학)	0.0488
3	관광학 (사회과학)	경영학 (사회과학)	0.0472
4	경제학 (사회과학)	경영학 (사회과학)	0.0438
5	심리과학 (사회과학)	기타사회과학 (사회과학)	0.0434
6	사회복지학 (사회과학)	기타사회과학 (사회과학)	0.0408
7	한국어와문학 (인문학)	기타인문학 (인문학)	0.0358
8	기타사회과학 (사회과학)	생활과학 (자연과학)	0.0352
9	무역학 (사회과학)	경영학 (사회과학)	0.0347
10	디자인 (예술체육)	예술일반 (예술체육)	0.0328

분석 결과, 가장 유사하다고 판단된 학문 분야들은 영어와문학(인문학) 분야와 언어학(인문학) 분야로 유사도 0.0540 으로 나타났다. 자카드 유사 계수는 두 분야의 성분이 완벽히 일치하는 경우 1 의 값을, 일치하는 성분이 전혀 없는 경우 0 의 값을 가진다. 이러한 특성을 고려할 때, 각 학문 분야에서 주로 인용하는 참고문헌이 다른 학문 분야와 공유되는 경우가 드물어 학문 분야 간 일치하는 참고문헌의 수는 비교적 적은 것으로 나타났다.

한편, 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 9. 자카드 유사 계수를 활용한 상위 대분류 영역이 서로 다른

학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
8	기타사회과학 (사회과학)	생활과학 (자연과학)	0.0352
11	사회복지학 (사회과학)	생활과학 (자연과학)	0.0295
12	교육학 (사회과학)	생활과학 (자연과학)	0.0292
17	식품과학 (농수해양)	생활과학 (자연과학)	0.0281
21	관광학 (사회과학)	체육 (예술체육)	0.0271
22	심리과학 (사회과학)	생활과학 (자연과학)	0.0259
34	식품과학 (농수해양)	기타자연과학 (자연과학)	0.0214
36	심리과학 (사회과학)	여성학 (복합학)	0.0211
39	심리과학 (사회과학)	정신과학 (의약학)	0.0208
40	공학일반 (공학)	지역개발 (사회과학)	0.0208

분석 결과, 빈도 지수를 활용한 유사 지수 결과와는 달리 다양한 학문 분야들이 상위 항목에 나타났다. 이러한 결과의 차이는 유사 지수 도출 방법의 특성의 차이에서 기인한 것으로 판단된다. 학문 분야에 속한 문헌이 많을수록 다른 학문 분야와의 공통 참고문헌이 발생할 확률이 높다. 그런데 빈도 지수 기반 분석에서는 공통된 참고문헌의 빈도만이 가중치로 작용하기 때문에 문헌이 많이 속한 학문 분야가 항목의 상위에 나타날 가능성이 높다. 그러나 유사도 지수 기반 분석에서는 공통 참고문헌이 전체 참고문헌에서 차지하는 비율 또한 유사 지수에 영향을 주기 때문에, 분야에 속한 문헌의 수에 의한 단순 영향이 상당 부분 제거되어 실제로 유사성이 높은 학문 분야 간 관계들이 항목의 상위에 나타날 가능성이 높다.

자카드 유사 계수를 활용한 학문 분야 간 관계의 시각화 결과는 다음과 같다.

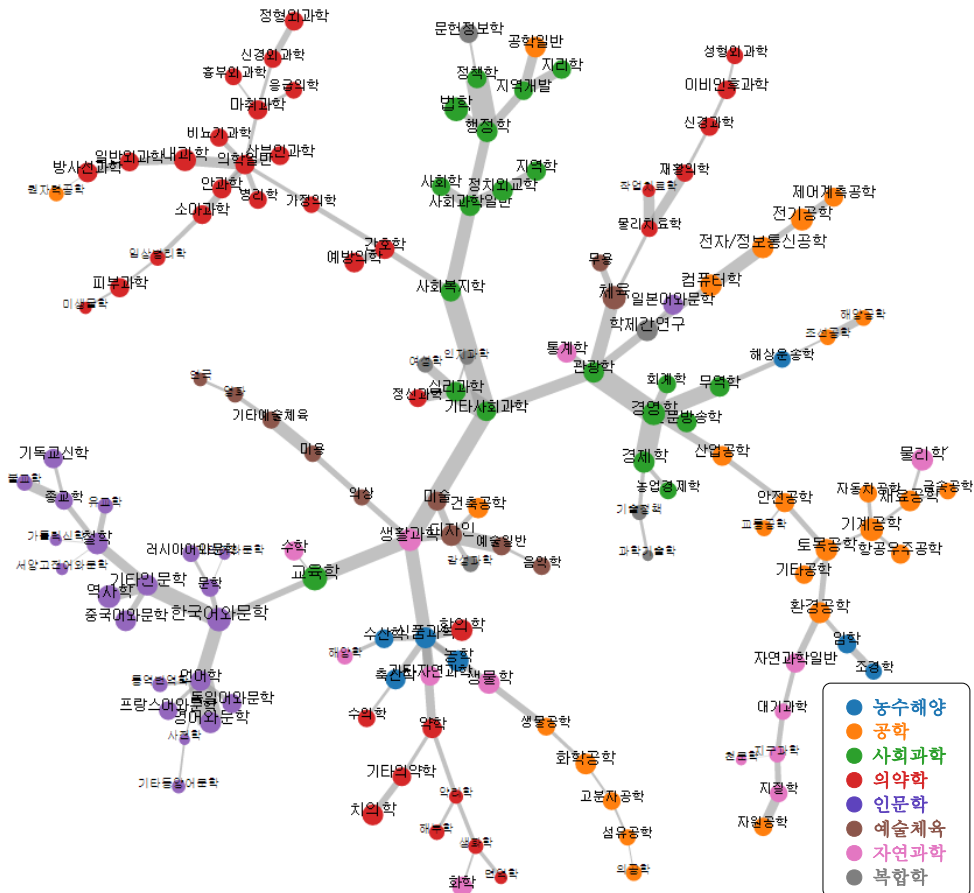


그림 4. 자카드 유사 계수를 활용한 학문 분야 간 패스파인더 네트워크

시각화 결과, 인문학 영역의 경우 빈도 지수 기반 시각화 결과와 마찬가지로 동일 대분류 영역 내의 분야 간 결합이 상대적으로 강하게 나타났으나, 사회과학 영역의 경우 정치/행정 관련 분야와 경영/경제 관련 분야가 분리되어 나타나는 경향을 보였다. 의약학 영역 역시 의약과 약학 관련 분야, 재활의학 관련 분야가 분리되어 나타났으며, 공학 영역 또한 전자/전기/제어/컴퓨터학 관련 분야와

기계/재료/토목공학 관련 분야, 화학/생물공학 관련 분야, 조선/해양공학 관련 분야가 분리되어 나타났다. 예술체육 영역의 경우 생활과학 분야에 연계된 예술 관련 분야와 관광학 분야에 연계된 체육 관련 분야로 나뉘어 나타났다. 요약하자면, 시각화 결과의 전체적인 구조는 빈도 지수 기반의 시각화 결과들과 형태적인 차이가 있으며, 대분류 영역 내 관계가 강하게 나타났던 빈도 지수 기반 결과보다 실제 주제가 비슷한 영역 간 관계가 상대적으로 강하게 나타났다.

5.2.2 코사인 유사도 활용 결과

코사인 유사도를 활용한 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 10. 코사인 유사도를 활용한 학문 분야 간 유사 지수
(상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	철학 (인문학)	유교학 (인문학)	0.422
2	관광학 (사회과학)	경영학 (사회과학)	0.382
3	역사학 (인문학)	기타인문학 (인문학)	0.364
4	영어와문학 (인문학)	언어학 (인문학)	0.357
5	무역학 (사회과학)	경영학 (사회과학)	0.312
6	철학 (인문학)	기타인문학 (인문학)	0.262
7	정책학 (사회과학)	행정학 (사회과학)	0.260
8	한국어와문학 (인문학)	기타인문학 (인문학)	0.244
9	경제학 (사회과학)	경영학 (사회과학)	0.242
10	심리과학 (사회과학)	기타사회과학 (사회과학)	0.237

분석 결과, 철학(인문학) 분야와 유교학(인문학) 분야의 유사도가 0.422 로 가장 높게 나타났다. 코사인 유사도의 경우 두 영역의 성분이 완전히 반대되는 경우 -1 의 값을, 관계가 없는 경우 0 의 값을, 완전히 같은 경우 1 의 값을 가질 수 있다. 그러나 인용 분석의 경우 인용의 발생 특성상 음의 빈도를 가질 수 없으므로, 인용 분석에서의 코사인 유사도는 0 과 1 사이의 값을 가지게 된다. 따라서 상위 항목에 나타난 학문 분야들은 서로 어느 정도 유사한 관계가 있음을 알 수 있다.

위 분석 결과를 앞서 분석한 자카드 유사 계수 활용 결과와 비교하였을 때, 자카드 유사 계수보다 코사인 유사도가 상대적으로 높게 나타남을 확인할 수 있다. 이는 자카드 유사 계수가 단순히 성분의 유무만을 판단하여 유사도를 측정하는 데 비하여, 코사인 유사도의 경우 성분의 크기 또한 분석에 포함되기 때문이며, 특정 공통 문헌의 인용 횟수가 높은 경우 자카드 유사 계수보다 유사도가 높게 나타날 수 있다. 또한, 코사인 유사도 상위 항목에 인문학 영역이 상대적으로 많이 나타난 것은 발표 후 시간이 많이 흐른 문헌의 인용 빈도가 타 영역보다 높아 성분의 크기가 비교적 큰 인문학 영역의 고유 특성에 기인하였다고 해석할 수 있다.

한편, 상위 대분류 영역이 서로 다른 학문 분야 간 유사 지수 상위 10 항목은 다음과 같다.

표 11. 코사인 유사도를 활용한 상위 대분류 영역이 서로 다른
학문 분야 간 유사 지수 (상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
15	관광학 (사회과학)	체육 (예술체육)	0.185
16	기타사회과학 (사회과학)	생활과학 (자연과학)	0.184
21	심리과학 (사회과학)	여성학 (복합학)	0.153
22	경영학 (사회과학)	체육 (예술체육)	0.149
23	산업공학 (공학)	경영학 (사회과학)	0.147
24	역사학 (인문학)	미술 (예술체육)	0.146
28	심리과학 (사회과학)	정신과학 (의약학)	0.143
31	교육학 (사회과학)	생활과학 (자연과학)	0.141
32	생물공학 (공학)	생물학 (자연과학)	0.140
34	식품과학 (농수해양)	생활과학 (자연과학)	0.136

분석 결과, 상위 영역이 서로 다른 학문 분야 간 유사 관계는 15 위에서 처음으로 나타나 다른 방법에 비해 하위 순위에서 나타남을 확인할 수 있다. 연구 주제가 유사할 것으로 직관적으로 추정되는 학문 분야들의 관계가 상위 항목에 다수 나타났으며, 유사 지수는 최대 0.185로서 비교적 낮은 유사 관계가 있는 것으로 분석되었다.

코사인 유사도를 활용한 학문 분야 간 관계의 시각화 결과는 다음과 같다.

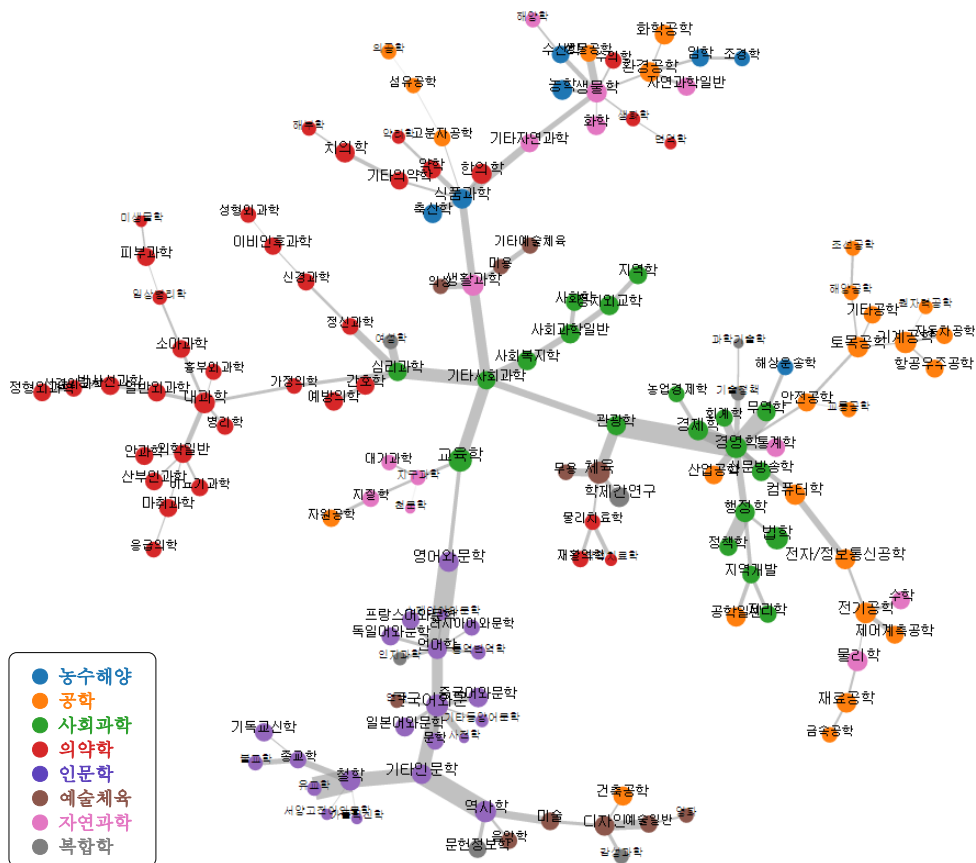


그림 5. 코사인 유사도를 활용한 학문 분야 간 패스파인더 네트워크

시각화 결과의 전체적인 형태는 자카드 유사 계수 활용 시와 비슷하게 나타났다. 코사인 유사도를 활용한 시각화 결과는 자카드 유사 계수 활용 결과와 비교하여 세부 학문 분야들의 위치와 관계 부분에서는 약간의 차이가 있으나, 실제 주제가 유사한 학문 영역 간 군집이 이루어지는 경향은 비슷하다.

5.2.3 피어슨 상관계수 활용 결과

피어슨 상관계수를 활용한 학문 분야 간 관계 지수 상위 10 항목은 다음과 같다.

표 12. 피어슨 상관계수를 활용한 학문 분야 간 유사 지수
(상위 10 항목)

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	문학 (인문학)	연극 (예술체육)	0.963
2	독일어와문학 (인문학)	프랑스어와문학 (인문학)	0.958
3	러시아어와문학 (인문학)	프랑스어와문학 (인문학)	0.958
4	통역번역학 (인문학)	독일어와문학 (인문학)	0.951
5	회계학 (사회과학)	무역학 (사회과학)	0.951
6	산업공학 (공학)	경제학 (사회과학)	0.950
7	기타동양어문학 (인문학)	일본어와문학 (인문학)	0.948
8	중국어와문학 (인문학)	기타동양어문학 (인문학)	0.946
9	산업공학 (공학)	무역학 (사회과학)	0.942
10	무역학 (사회과학)	기술정책 (복합학)	0.942

피어슨 상관계수의 경우 그 결과로서 -1 과 1 사이의 값을 가질 수 있으나, 코사인 유사도와 마찬가지로 인용의 발생 특성상 음의 빈도를 가질 수 없으므로, 인용 분석에서의 피어슨 상관계수는 0 과 1 사이의 값을 가지게 된다. 실제로 피어슨 상관계수의 도출 결과는 최댓값 0.963, 최솟값 0.216 으로 나타났다.

분석 결과, 앞서 도출한 자카드 유사 계수와 코사인 유사도보다 학문 분야 간 관계성이 매우 높게 나타났다. 일반적으로 분석 대상 간 피어슨 상관계수가 0.9 이상인 경우 분석 대상들은 서로 매우 높은 상관관계가 있다고 해석된다. 따라서 상위 항목에 나타난 학문 영역들은 서로 밀접한 관계가 있는 것으로 평가된다.

한편, 상위 대분류 영역이 서로 다른 학문 분야 간 관계 지수 상위 10 항목은 다음과 같다.

**표 13. 피어슨 상관계수를 활용한 상위 대분류 영역이 서로 다른
학문 분야 간 관계 지수 (상위 10 항목)**

등위	유사 학문 분야 (상위 대분류 영역)		지수
1	문학 (인문학)	연극 (예술체육)	0.963
6	산업공학 (공학)	경제학 (사회과학)	0.950
9	산업공학 (공학)	무역학 (사회과학)	0.942
10	무역학 (사회과학)	기술정책 (복합학)	0.942
13	회계학 (사회과학)	기술정책 (복합학)	0.936
14	산업공학 (공학)	회계학 (사회과학)	0.932
16	중국어와문학 (인문학)	연극 (예술체육)	0.931
20	정신과학 (의약학)	여성학 (복합학)	0.924
27	산업공학 (공학)	기술정책 (복합학)	0.911
32	기독교신학 (인문학)	음악학 (예술체육)	0.905

분석 결과, 상위 영역이 서로 다른 학문 분야 간 관계 지수도 매우 높은 수준으로 나타났다.

피어슨 상관계수를 활용한 학문 분야 간 관계의 시각화 결과는 다음과 같다.

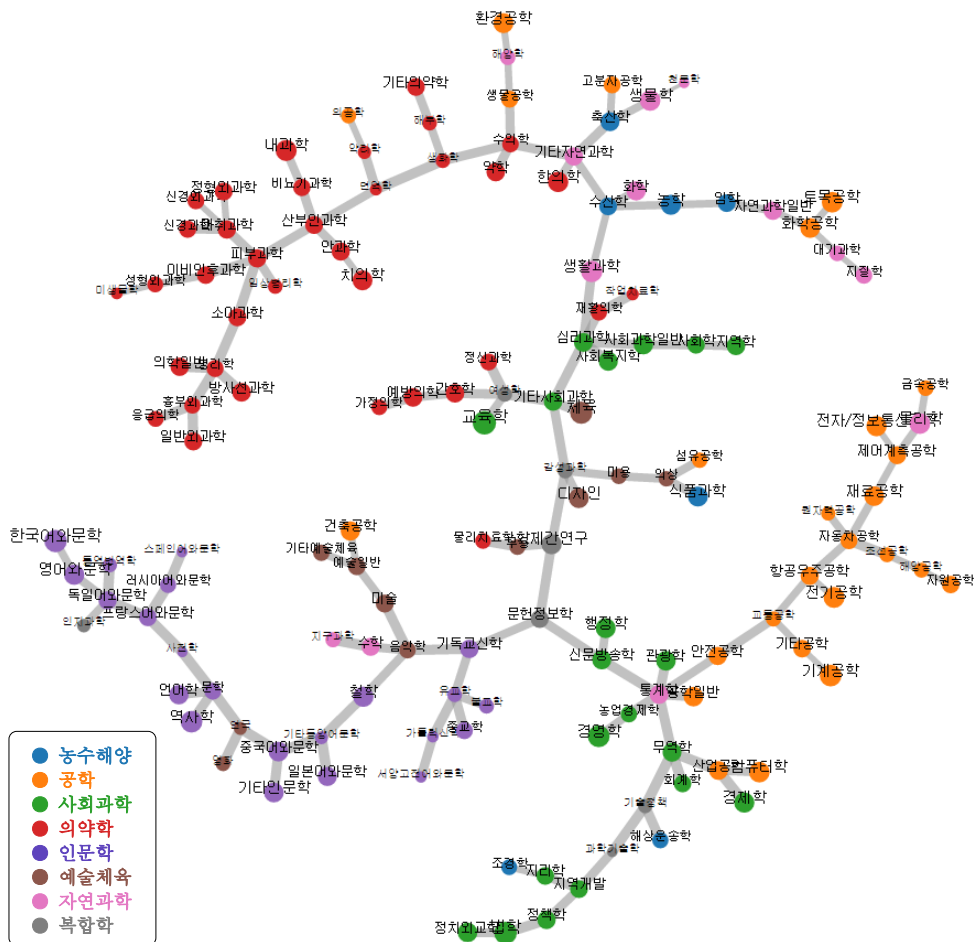


그림 6. 피어슨 상관계수를 활용한 학문 분야 간 패스파인더 네트워크

시각화 결과, 부분적인 군집의 형태는 어느 정도 관별할 수 있으나 그 계층 구조를 명확히 구분하기는 어려운 형태로 나타났다. 또한, 학문 영역 간 유사성이 나타났으나 그 관계를 직관적으로 이해하기 어려운 연결이 상당수 발견되었다. 피어슨 상관계수는 유의수준 0.01의 데이터만을 활용하였으므로 학문 분야 간 유사 관계를 분석하기 위한 유용한 정보를 제공할 수 있으나, 이를 활용한 시각화

결과는 학문 분야의 계층 구조를 분석하기 위해서는 적절하지 않은 방법으로 판단된다.

5.2.4 유사도 지수 기반 결과의 상호 비교

위에서 살펴본 세 가지 방법으로 도출한 유사도 지수 기반 결과의 공통된 특징들은 다음과 같이 나타났다.

1. 세 가지 방법 모두 학문 분야 간 관계를 시각화하는 것이 가능하였으며, 비슷한 주제를 가지는 군집을 판별할 수 있었다.
2. KCI 대분류 영역 중 인문학 영역과 의학 영역의 세부 학문 분야들은 동일 영역에 속한 학문 분야들과의 관계가 다른 영역에 비하여 세 가지 방법에서 모두 상대적으로 높게 나타났다.

한편, 선택한 방법에 따른 결과의 차이점 또한 존재하였다.

1. 자카드 유사 계수와 코사인 유사도를 활용한 경우 전체적인 구조는 비슷하게 나타났으나, 세부 구조는 몇몇 학문 분야들의 위치와 관계 측면에서 약간의 차이를 보였다.
2. 코사인 유사도의 경우 자카드 유사도보다 영역 간 유사성이 상대적으로 크게 나타났으며, 피어슨 상관계수의 경우 다른 두 경우보다 더 크게 나타났다.
3. 피어슨 상관계수를 활용한 경우, 다른 방법들을 활용한 경우에 비하여 시각화 결과에 대한 직관적인 관계 구조 파악이 상대적으로 어려웠으며, 동일 대분류 영역에 속하는 학문 영역 간 군집 경향은 어느 정도 나타났으나 그 연결 형태는 분석하기 어려운 경우가 다수 발견되었다.

위의 각 방법에 따른 도출 결과의 공통점과 차이점을 바탕으로, 유사도 지수를 통하여 학문 영역 간 관계를 분석하여 시각화하는 경우

자카드 유사 계수 또는 코사인 유사도를 활용하는 것이 피어슨 상관계수를 활용하는 것에 비해 그 구조를 분석하는 데 유리하다고 평가할 수 있다. 이는 사회과학 및 자연과학 영역에서의 위의 각 방법으로 도출된 지수를 활용한 시각화 결과에 대하여 해석하였던 Boyack의 연구와도 합치하는 결과이다.

5.3 활용 지수에 따른 시각화 결과 비교

위에서 살펴본 크게 두 분류의 유사 지수 도출 방법에 따른 시각화 결과들은 다음과 같은 차이점이 발견되었다.

1. 빈도 기반 방법을 활용하여 시각화한 경우 다른 교점에 비해 연결도가 상당히 높은 허브(hub) 교점들이 발견되었으나, 유사도 기반 방법을 활용하여 시각화한 경우에 나타난 허브 교점들은 이전의 경우에 비해 다른 교점과의 연결도의 차이가 작게 나타났다. 이러한 차이는 유사 지수 도출 방법의 차이에 의하여 발생한 것으로 판단할 수 있다. 빈도 기반 방법의 경우, 분석 대상 분야에 속한 문헌의 증가는 공통 참고문헌의 발생 및 그 인용횟수의 증가를 가져올 가능성이 높다. 따라서 상대적으로 큰 수의 문헌을 가진 학문 분야가 다른 학문 분야들과 관계를 맺을 확률이 높고, 이러한 효과는 허브 교점의 연결도가 높아지는 방향으로 작용한다. 그러나 유사도 기반 방법의 경우 분석 대상 분야에 속한 문헌의 증가가 유사 지수의 증가를 보장하지 않으며, 실질적인 참고문헌의 중복 비율이 유사 지수에 영향을 주게 되므로 단순히 많은 문헌이 속한 학문 분야에 집중적으로 연결이 형성되는 모습은 나타나지 않으며, 허브 교점의 연결도는 상대적으로 낮아지게 된다.

2. 빈도 기반 방법을 활용한 경우 교점 간 계층 구조가 유사도 기반 방법을 활용한 것에 비하여 상대적으로 명확하게 나타났다. 이는 동시인용분석 방법을 활용하여 원래의 빈도 행렬과 피어슨 상관계수 행렬의 패스파인더 네트워크 알고리즘 시각화 결과를 비교하고, 원래의 빈도 행렬을 활용하는 것이 적합하다고 판단하였던 White 의 결과와도 어긋나지 않는다. 그러나 비슷한 주제를 가지는 군집을 판별하기 위해서는 유사도 기반 방법을 활용하는 것이 빈도 기반 방법을 활용하는 것에 비해 더 적절하다고 판단되었다. 이는 빈도 기반 방법을 활용하여 결과를 도출하여 시각화한 경우 강한 연결도를 가지는 허브 교점으로 인하여 세부 학문 군집의 구분이 어렵지만, 유사도 기반 방법을 활용하는 경우는 연구 주제가 유사한 학문 분야별로 군집이 이루어져 그 구분이 수월하기 때문이다.

위의 유사 지수 도출 방법의 특성에 따라 나타난 차이점을 바탕으로 각 방법의 활용 방안을 구분하면, 전통적인 계층 구조의 학문 분류를 의도하는 경우 빈도 기반 방법을 활용하는 것이 분석에 유리하며, 주제에 기반을 둔 학문 분야 군집화를 의도하는 경우 유사도 기반 방법을 활용하는 것이 유리하다고 평가할 수 있다.

5.4 연구의 한계

본 연구를 통해 위와 같은 학문 분야 간 유사 관계 도출 및 시각화에 대한 결과가 도출되었지만, 한편으로는 동일 참고문헌의 판별, KCI 참고문헌의 미비, 그리고 학문 영역 간 관계 평가에 대한 어려움 등 연구의 한계점 또한 존재하였다.

5.4.1 동일 참고문헌의 판별 문제

문헌 정보의 전처리 과정에서, 참고문헌 목록 중에 동일 문헌으로 의심되거나 서로 다른 고유 코드가 부여된 경우를 찾을 수 있었다. 이는 인용 정보 기록 단계에서 학문 영역 및 개인별 인용 정보 기록 방식의 차이 때문에 발생하는 문제이다.

참고문헌이 KCI 등재 문헌인 경우 일차적으로 입력 과정에서의 검색을 통하여 동일 문헌이 판단되어 등록된 경우도 있었다. 그러나 KCI 비등재 문헌의 경우 참고문헌 정보의 오류 여부를 판단할 수 없는 경우가 대부분이기 때문에 문헌에 기록된 그대로 입력된 경우가 많으며, 이는 서지결합분석 과정에서 참고문헌의 동일성을 판단하는데 장애 요소로 작용하게 된다. 본 연구에서는 이름 및 연도, 제목을 식별자로 활용하여 SHA1 해시 코드를 추출하고, 이를 비교하여 참고문헌의 동일성 여부 판단을 통해 이러한 참고문헌 정보 표현상의 차이로 발생하는 문제를 최소화하기 위해 노력하였다. 그러나 이러한 노력에도 불구하고 분석 대상 중 같은 문헌으로 추정되는 참고문헌의 경우가 종종 발견되었다. 같은 문헌으로 추정되나 동일 문헌으로 처리되지 못한 경우는 중간 이름 또는 첫 글자 표현 등으로 인한 저자명 표현 방식의 차이에 의해 발생하거나, 부제의 생략으로 인한 제목 표현 방식의 차이에 의해 발생하는 경우가 많았다.

분석 대상인 참고문헌의 수가 현재 약 1,200 만 건으로 상당히 많고 그 수는 등재 문헌이 증가함에 따라 지속해서 증가할 것으로 예측된다. 따라서 수작업을 통한 인용 정보의 정제보다는 인용 목록에 대한 형태소 분석 및 유사도 수준에 따른 묶음 설정 등의 기술적 정제 과정을 통하여 동일 문헌을 판단하여 처리한다면 빠르고 정확한 인용

정보의 처리가 가능할 것으로 생각된다. 이러한 동일 참고문헌 판별 방법의 적용을 통하여 참고문헌의 중복 여부를 확실하게 판단한다면, 학문 영역 간 유사성 또한 더욱 명확하게 분석할 수 있을 것으로 생각된다.

5.4.2 KCI 참고문헌 데이터의 미비

KCI DB 현황의 연도별 데이터 현황에 따르면, 2011 년까지의 등재 문헌과 이에 대한 참고문헌 수는 지속해서 증가하였다. 그러나 2012 년의 경우, 등재 문헌 수는 증가하였으나 이에 대한 참고문헌은 처리되지 않아 6,241 건(전년 대비 0.3%)에 불과하며, 2013 년 또한 당해 년도 문헌 정보 수집 및 처리에 필요한 필연적인 시간 지연으로 인하여 등재 문헌 수 및 이에 대한 참고문헌의 수가 각각 67,719 건과 2,385 건으로 타 연도 대비 매우 부족함을 확인할 수 있었다. 이러한 문제는 수집된 문헌의 전처리 과정에서도 상당 부분 확인되었는데, 2012 년 또는 2013 년 문헌의 상당수가 실제로는 참고문헌이 존재함에도 불구하고 참고문헌 수가 0 으로 등록된 경우를 다수 발견할 수 있었다.

이러한 최근 문헌에 대한 참고문헌 정보의 미비 문제는 비교적 최근에 생성된 학문 분야들의 빈도 기반 관계 측정 시 다른 전통적인 학문 분야들이 가지는 관계에 비해 상대적으로 약하게 측정되게 하는 문제로 작용하였다. 특히 영역에 속한 문헌의 수가 상대적으로 적은 신생 학문 분야의 경우 특정 학술지의 문헌에 대한 참고문헌이 먼저 처리됨으로 인하여 그 관계가 상대적으로 치우쳐 나타나는 현상을 관찰할 수 있었다. 또한, 동시인용분석보다 상대적으로 최근의 학문 간 관계를 알아보기 위하여 서지결합분석 방법을 선택하였지만, 기초

데이터의 미비함으로 인하여 최소 2 년 전의 학문 간 관계만을 분석할 수 있게 되어 그 결과적 측면의 한계가 발생하였다.

5.4.3 도출된 관계에 대한 평가의 어려움

참고문헌의 관계 분석을 통하여 학문 간 관계를 도출하였지만, 도출된 관계를 평가할 수 있는 적절한 방법을 탐색하기 어려운 것이 한계로 존재하였다. 관계의 검증을 위한 객관적인 방법이 존재하지 않기 때문에, 본 연구와 비슷한 연구를 수행한 선행 연구의 경우 결과에 대해 평가를 하지 않거나, 연구자 자신이 속한 연구 분야에 대하여만 평가하거나, 기존에 존재하는 학문 분류 체계와의 일치 정도를 통하여 그 적합성을 평가하는 경우가 있었다. 그러나 연구자 자신이 속한 연구 분야에 대하여만 평가하는 경우 다른 영역에 대한 평가는 이루어지지 않는 문제가 있으며, 기존에 존재하는 학문 분류 체계와의 일치 정도로 적합성을 평가하는 경우 기존 분류 체계가 학문 영역 간 관계 구조를 적절히 표현하고 있는지 판단하기 어렵다. 또한, 어떠한 학문 분류 체계가 비교의 대상으로 적절하여 옳은 기준임을 증명하는 것이 불가능하며, 만약 비교의 대상으로 적절한 기준이 존재한다면 다른 방법으로 학문의 분류 및 학문 간 관계를 생성할 필요성이 사라지는 문제가 발생한다. 따라서 학문 영역 간 관계 도출 결과에 대한 검증을 위해서는 연구 분야의 전문가 면담 또는 문헌 간 연계의 시간적 추이 변화 등을 활용한 반복적인 관계 평가 과정을 통하여 도출된 관계의 적합성 여부를 판단해야 할 것으로 생각된다.

제 6 장 결론

본 연구에서는 서지결합분석을 활용하여 국내 학문 영역에 대한 영역 간 관계를 다양한 유사 지수를 통해 도출하고, 이에 기초한 광역 과학 지도를 생성하여 그 결과들을 서로 비교하였다. KCI 의 중분류 항목 총 151 개 항목을 분석 대상인 학문 영역으로 간주하여, 등재된 문헌 약 93 만 건에 나타난 공통 참고문헌을 바탕으로 그들의 인용 관계에 기반을 둔 학문 영역 간 관계를 분석하고 시각화하였다.

영역 간 유사 관계를 측정하는 데는 총 6 가지 방법을 활용하였으며, 크게 빈도 기반 방법과 유사도 기반 방법으로 분류하였다. 빈도 기반 방법으로는 단순 방법과 최소 방법, 혼합 방법을 적용하였으며, 혼합 방법보다 단순 방법 또는 최소 방법을 활용한 경우 결과를 해석하기에 수월함을 확인할 수 있었다. 유사도 기반 방법으로는 자카드 유사 계수와 코사인 유사도, 피어슨 상관계수를 적용하였으며, 피어슨 상관계수보다 자카드 유사 계수 또는 코사인 유사도를 활용한 경우 결과를 해석하기에 수월함을 확인할 수 있었다.

시각화 과정에서는 패스파인더 네트워크 알고리즘을 적용하여 관계를 직관적으로 판단할 수 있도록 하였다. 또한, 시각화는 D3.js 라이브러리를 활용하여 웹 브라우저에서 접근 가능한 형태로 구현하였다. 따라서 특별한 작업 없이도 본 연구의 결과를 인터넷을 통해 자유롭게 접근할 수 있게 하였다. 이를 통하여 연구의 결과가 단순히 학술적인 의미로만 해석되고 그치는 것이 아니라, 학계 또는 연구기관에 속한 연구자들이 자신의 관심 학문 영역에 대한 지적 구조 및 위치를 확인하고, 자신의 연구에 활용 가능한 관계 정보를 획득할 가능성을 열어둘 수 있다.

본 연구의 최종 결과물로서, 학문 영역 간 관계가 시각적으로 표현되어 그 구조를 직관적으로 파악할 수 있는 광역 과학 지도를 생성하였다. 또한, 기존의 고전적인 학문 분류와는 다른 형태를 보이는, 실제 관련성이 높은 학문 영역들의 군집 및 구분을 시도하였다. 이러한 광역 과학 지도의 도출은 연구자들에게 자신의 연구 영역과 인접해 있는 다른 영역에 대한 관심을 불러일으켜 학문의 세분화에 따른 연구 영역의 단편화 현상을 경감시키고, 나아가 다른 영역의 연구자들과의 공동 연구 및 학제적 연구의 시도 가능성을 증가시킬 수 있을 것으로 기대된다.

참고문헌

- [1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge University Press, 1994.
- [2] B. Cronin, *The Citation Process: The Role and Significance of Citations in Scientific Communication*. Taylor Graham, 1984.
- [3] H. G. Small, “Co-Citation in the Scientific Literature: A New Measure of the Relationship between Two Documents,” *Journal of the American Society for Information Science*, vol. 24, no. 4, pp. 265–269, 1973.
- [4] H. D. White and K. W. McCain, “Bibliometrics,” *Annual Review of Information Science and Technology*, vol. 24, pp. 119–186, 1989.
- [5] H. G. Small and B. C. Griffith, “The Structure of Scientific Literatures I: Identifying and Graphing Specialties,” *Science Studies*, vol. 4, no. 1, pp. 17–40, 1974.
- [6] H. G. Small, “A Co-Citation Model of a Scientific Specialty: A Longitudinal Study of Collagen Research,” *Social Studies of Science*, vol. 7, no. 2, pp. 139–166, 1977.
- [7] H. G. Small, “Paradigms, Citations, and Maps of Science: A Personal History,” *Journal of the American Society for Information Science and Technology*, vol. 54, no. 5, pp. 394–399, 2003.
- [8] M. S. Granovetter, “The Strength of Weak Ties,” *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

- [9] F. J. González, C. B. Castro, J. C. C. Bueno, and J. L. G. González, “Dominant approaches in the field of management,” *International Journal of Organizational Analysis*, vol. 9, no. 4, pp. 327–353, 2001.
- [10] H. D. White and B. C. Griffith, “Author cocitation: A literature measure of intellectual structure,” *Journal of the American Society for Information Science*, vol. 32, no. 3, pp. 163–171, 1981.
- [11] H. D. White, “A Cocitation Map of The Social Indicators Movement,” *Journal of the American Society for Information Science*, vol. 34, no. 5, pp. 307–312, 1983.
- [12] H. D. White, “Author cocitation analysis and Pearson’s r ,” *Journal of the American Society for Information Science and Technology*, vol. 54, no. 13, pp. 1250–1259, 2003.
- [13] R. Rousseau and A. Zuccala, “A Classification of Author Co-citations: Definitions and Search Strategies.,” *Journal of the American Society for Information Science and Technology*, vol. 55, no. 6, pp. 513–529, 2004.
- [14] K. W. McCain, “The author cocitation structure of macroeconomics,” *Scientometrics*, vol. 5, no. 5, pp. 277–289, 1983.
- [15] H. D. White, *Author co-citation analysis: Overview and defense*. Newbury Park: SAGE, 1990, pp. 84–106.
- [16] 이재윤, “서지적 저자결합분석: 연구동향 분석을 위한 새로운 접근,” *한국정보관리학회*, vol. 25, no. 1, pp. 173–190, 2008.

- [17] D. Zhao and A. Strotmann, "Author bibliographic coupling: Another approach to citation-based author knowledge network analysis," *Proceedings of the American Society for Information Science and Technology*, vol. 45, no. 1, pp. 1–10, 2008.
- [18] K. W. McCain, "Longitudinal author cocitation mapping: The changing structure of macroeconomics," *Journal of the American Society for Information Science*, vol. 35, no. 6, pp. 351–359, 1984.
- [19] K. W. McCain, "Cocited author mapping as a valid representation of intellectual structure," *Journal of the American Society for Information Science*, vol. 37, no. 3, pp. 111–122, 1986.
- [20] K. W. McCain, "Mapping Authors in Intellectual Space: A Technical Overview," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 433–443, 1990.
- [21] F. J. Acedo, C. BarrosoCasanueva, C. C. Rocha, and J. L. Galán, "Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis," *Journal of Management Studies*, vol. 43, no. 5, pp. 957–983, 2006.
- [22] P. Ahlgren, B. Jarneving, and R. Rousseau, "Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 6, pp. 550–560, 2003.
- [23] H. D. White, "Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists,"

Journal of the American Society for Information Science and Technology, vol. 54, no. 5, pp. 423–434, 2003.

- [24] H. Kreuzman, “A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science,” *Scientometrics*, vol. 50, no. 3, pp. 525–539, 2001.
- [25] J. E. Andrews, “An author co-citation analysis of medical informatics,” *Journal of the Medical Library Association*, vol. 91, no. 1, pp. 47–56, 2003.
- [26] 김희정, “저자 동시인용분석에 의한 국내외 기록관리학 분야의 지적구조 비교에 관한 연구,” *한국문헌정보학회*, vol. 39, no. 3, pp. 207–224, 2005.
- [27] K. W. McCain, “Gene’s worlds: Comparing Eugene Garfield’s co-citation and tri-citation image PFNets over time,” *Proceedings of the American Society for Information Science and Technology*, vol. 45, no. 1, pp. 1–9, 2008.
- [28] N. C. Mullins, L. L. Hargens, P. K. Hecht, and E. L. Kick, “The Group Structure of Cocitation Clusters: A Comparative Study,” *American Sociological Review*, vol. 42, no. 4, pp. 552–562, 1977.
- [29] H. D. White, “Introduction,” *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 430–432, 1990.
- [30] R. E. Rice, C. L. Borgman, and B. Reeves, “Citation networks of communication journals, 1977–1985: Cliques and positions, citations made and citations received,” *Human Communication Research*, vol. 15, no. 2, pp. 256–283, 1988.

- [31] K. W. McCain, "Mapping Authors in Intellectual Space: Population Genetics in the 1980s," *Communication Research*, vol. 16, no. 5, pp. 667-681, 1989.
- [32] W. Paisley, "An Oasis Where Many Trails Cross: The Improbable Cocitation Networks of a Multidiscipline," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 459-468, 1990.
- [33] S. P. Nerur, A. A. Rasheed, and V. Natarajan, "The intellectual structure of the strategic management field: an author co-citation analysis," *Strategic Management Journal*, vol. 29, no. 3, pp. 319-336, 2008.
- [34] 김영진, "논문의 동시인용을 통한 지적(知的)구조의 규명에 관한 연구," 정보관리학회지, vol. 3, no. 1, pp. 103-135, 1986.
- [35] 민윤경, "저자동시인용분석에 의한 경영학 분야의 지적 구조 규명에 관한 연구," 이화여자대학교 대학원, 1992.
- [36] 김도미, "저자동시인용분석과 인용한 문헌의 색인어 분석에 의한 지적 구조의 규명," 정보관리연구, vol. 1, no. 24, pp. 32-57, 1993.
- [37] M. M. Kessler, "Bibliographic coupling between scientific papers," *American Documentation*, vol. 14, no. 1, pp. 10-25, 1963.
- [38] J. Qiu, "Methods of citation analysis," *Informetrics*, Wuhan: Wuhan University Press, 2007.
- [39] B. H. Weinberg, "Bibliographic Coupling: A Review," *Information Storage and Retrieval*, vol. 10, no. 5, pp. 189-196, 1974.

- [40] L. Egghe and R. Rousseau, "Co-citation, bibliographic coupling and a characterization of lattice citation networks," *Scientometrics*, vol. 55, no. 3, pp. 349–361, 2002.
- [41] D. Zhao and A. Strotmann, "Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 13, pp. 2070–2086, 2008.
- [42] R. Ma, "Author bibliographic coupling analysis: A test based on a Chinese academic database," *Journal of Informetrics*, vol. 6, no. 4, pp. 532–542, 2012.
- [43] R. Rousseau, "Bibliographic coupling and co-citation as dual notions," *A Festschrift in honour of Peter Ingwersen, special volume of the e-zine of the ISSI 2010*, pp. 173–183, 2010.
- [44] W. Glänzel and H.-J. Czerwon, "A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level," *Scientometrics*, vol. 37, no. 2, pp. 195–221, 1996.
- [45] S. A. Morris, "Manifestation of emerging specialties in journal literature: A growth model of papers, references, exemplars, bibliographic coupling, cocitation, and clustering coefficient distribution," *Journal of the American Society for Information Science and Technology*, vol. 56, no. 12, pp. 1250–1273, 2005.
- [46] M.-H. Huang, L.-Y. Chiang, and D.-Z. Chen, "Constructing a patent citation map using bibliographic coupling: A study of

- Taiwan's high-tech companies," *Scientometrics*, vol. 58, no. 3, pp. 489–506, 2003.
- [47] O. Kuusi and M. Meyer, "Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm," *Scientometrics*, vol. 70, no. 3, pp. 759–777, 2007.
- [48] E. Bassecoulard, A. Lelu, and M. Zitt, "Mapping nanosciences by citation flows: A preliminary analysis," *Scientometrics*, vol. 70, no. 3, pp. 859–880, 2007.
- [49] B. Jarneving, "A comparison of two bibliometric methods for mapping of the research front," *Scientometrics*, vol. 65, no. 2, pp. 245–263, 2005.
- [50] B. Jarneving, "Bibliographic coupling and its application to research-front and other core documents," *Journal of Informetrics*, vol. 1, no. 4, pp. 287–307, 2007.
- [51] K. W. Boyack, K. Börner, and R. Klavans, "Mapping the structure and evolution of chemistry research," *Scientometrics*, vol. 79, no. 1, pp. 45–60, 2009.
- [52] S. A. Morris, G. Yen, Z. Wu, and B. Asnake, "Time line visualization of research fronts," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 5, pp. 413–422, 2003.
- [53] H. D. White, "Authors as citers over time," *Journal of the American Society for Information Science and Technology*, vol. 52, no. 2, pp. 87–108, 2001.

- [54] L. Leydesdorff, "BibCoupl.exe for Bibliographic Coupling among Authors," 2001. [Online]. Available: <http://www.leydesdorff.net/software/bibcoupl/>. [Accessed: 19-Dec-2013].
- [55] 김희전 and 조현양, "저자동시인용분석과 저자서지결합분석에 의한 지적 구조 분석: 사회복지학 분야를 중심으로," 정보관리학회지, vol. 3, no. 27, pp. 283-306, 2010.
- [56] 변지혜 and 정은경, "저자서지결합분석에 의한 국내 전기공학 분야 지적구조에 관한 연구," 정보관리연구, vol. 42, no. 4, pp. 75-94, 2011.
- [57] 김희진, "한국 미술치료학 분야의 지적구조 분석: 저자동시인용분석과 저자서지결합분석을 중심으로," 경성대학교 대학원, 2012.
- [58] 민형진, "국내 IT 아웃소싱 연구의 지적구조에 관한 연구: 저자서지결합분석을 중심으로," 건국대학교 대학원, 2012.
- [59] 박지연 and 정동열, "저자서지결합분석에 의한 문헌정보학의 지적구조 분석에 관한 연구," 정보관리학회지, vol. 30, no. 4, pp. 31-59, 2013.
- [60] R. J. Tijssen and A. F. Van Raan, "Mapping Changes in Science and Technology Bibliometric Co-Occurrence Analysis of the R&D Literature," *Evaluation Review*, vol. 18, no. 1, pp. 98-115, 1994.
- [61] L. Egghe, "Expansion of the field of informetrics: Origins and consequences," *Information Processing and Management*, vol. 41, no. 6, pp. 1311-1316, 2005.

- [62] L. Egghe, “Expansion of the field of informetrics: the second special issue,” *Information Processing and Management*, vol. 42, no. 6, pp. 1405–1407, 2006.
- [63] 이재윤, “지적 구조 규명을 위한 네트워크 형성 방식에 관한 연구,” *한국문헌정보학회지*, vol. 2, no. 40, pp. 333–355, 2006.
- [64] R. W. Schvaneveldt, *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ, USA: Ablex Publishing, 1990.
- [65] S. Noel, C.-H. Chu, and V. Raghvan, “Co-Citation count vs correlation for influence network visualization,” *Information Visualization*, vol. 2, no. 3, pp. 160–170, 2003.
- [66] Y. Ding, G. G. Chowdhury, and S. Foo, “Bibliometric Cartography of Information Retrieval Research by Using Co-word Analysis,” *Information Processing and Management*, vol. 37, no. 6, pp. 817–842, 2001.
- [67] K. Börner, “Studying the emergent ‘Global Brain’ in large-scale co-author networks and mapping the ‘Backbone of Science’,” 2005. [Online]. Available: <http://vw.indiana.edu/talks-spring05/borner.ppt>. [Accessed: 19-Dec-2013].
- [68] E. Garfield and R. K. Merton, *Citation indexing: Its theory and application in science, technology, and humanities*. New York: Wiley, 1979.
- [69] C. Chen, “Generalised similarity analysis and pathfinder network scaling,” *Interacting with Computers*, vol. 10, no. 2, pp. 107–128, 1998.

- [70] C. Chen, *Information Visualization: Beyond the Horizon*. Springer, 2004, p. 316.
- [71] K. Börner, C. Chen, and K. W. Boyack, “Visualizing knowledge domains,” *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 179–255, 2003.
- [72] C. Chen, *Mapping Scientific Frontiers: The Quest for Knowledge Visualisation*, vol. 59. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003, pp. 364–369.
- [73] F. Rossi, “Visualization methods for metric studies,” in *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics*, 2006, pp. 356–366.
- [74] L. S. Marion and K. W. McCain, “Contrasting views of software engineering journals: Author cocitation choices and indexer vocabulary assignments,” *Journal of the American Society for Information Science and Technology*, vol. 52, no. 4, pp. 297–308, 2001.
- [75] K. W. McCain, J. M. Verner, G. W. Hislop, W. Evanco, and V. Cole, “The use of bibliometric and knowledge elicitation techniques to map a knowledge domain: Software Engineering in the 1990s,” *Scientometrics*, vol. 65, no. 1, pp. 131–144, 2005.
- [76] K. W. McCain, “The structure of biotechnology R & D,” *Scientometrics*, vol. 32, no. 2, pp. 153–175, 1995.
- [77] C. Chen, “Visualising semantic spaces and author co-citation networks in digital libraries,” *Information Processing and Management*, vol. 35, no. 3, pp. 401–420, 1999.

- [78] C. Chen, "CiteSpace II: Detecting and Visualizing Emerging Trends," *Journal of the American Society for Information Science and Technology*, vol. 57, no. 3, pp. 359–377, 2006.
- [79] D. J. deSolla Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, pp. 510–515, 1965.
- [80] D. J. deSolla Price, "The science of scientist," *Medical Opinion and Review*, vol. 1, no. 10, pp. 88–97, 1966.
- [81] E. Garfield, "Scientography: Mapping the tracks of science," *Current Contents: Social & Behavioural Sciences*, vol. 7, no. 45, pp. 5–10, 1994.
- [82] B. Vargas-Quesada and F. de Moya-Anegón, *Visualizing the structure of science*. Springer, 2007.
- [83] H. G. Small, "Visualizing Science by Citation Mapping," *Journal of the American society for Information Science*, vol. 50, no. 9, pp. 799–813, 1999.
- [84] H. G. Small and E. Garfield, "The Geography of Science: Disciplinary and National Mappings," *Journal of information science*, vol. 11, no. 4, pp. 147–159, 1985.
- [85] F. de Moya-Anegón, B. Vargas-Quesada, V. Herrero-Solana, Z. Chinchilla-Rodríguez, E. Corera-Álvarez, and F. J. Munoz-Fernández, "A new technique for building maps of large scientific domains based on the cocitation of classes and categories," *Scientometrics*, vol. 61, no. 1, pp. 129–145, 2004.
- [86] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory social network analysis with Pajek*. Cambridge University Press, 2005.

- [87] 이재윤, “국내 광역 과학 지도 생성 연구,” 정보관리학회지, vol. 24, no. 3, pp. 363–383, 2007.
- [88] 홍종선, “KCI 기반 Kor-Factor (Korea Factor) 모형 개발 기초연구.” 2007.

Abstract

Analyze and visualize the relationship of domestic academic disciplines using bibliographic coupling

Gibeom Park

Program in Digital Contents and Information Studies

Department of Transdisciplinary Studies

The Graduate School

Seoul National University

This study derived the relationship of domestic academic disciplines by using bibliographic coupling which is one of the methods to analyze citation data, and visualized the science map with the relationship results. Documents and citation data of the journals which are listed on Korea Citation Index (KCI) were used for analyzing and visualizing the relationship of the disciplines. *Simple method*, *Minimum method* and *Combined method* were used for deriving the relationship based on frequencies of common references. Jaccard similarity coefficient, Cosine similarity, and Pearson correlation coefficient were used for deriving the relationship based on similarities between the disciplines.

Next, the relation matrices of the disciplines were calculated by above each methods, and the commons and differences were found by comparing the visualized results based on the matrices. Pathfinder network algorithm (PFNet) was applied to visualization process, which intended that links of highly related disciplines could be identified intuitively.

With the above process, the relationship of academic disciplines, which has been actively researched in last decade were derived by various quantitative methods. Then, science maps were generated by the derived results. At last, an appropriate method for visualizing the relationship of the disciplines was empirically suggested.

Keywords: bibliographic coupling, scientography, pathfinder network, information visualization, Korea Citation Index, KCI

Student Number: 2012-22463